ESSAYS ON VARIABLE ELASTICITY OF SUBSTITUTION, ECONOMIC GROWTH, AND HUMAN CAPITAL OUTCOMES

Sharmila Gamlath

Bsc (Hons), PGD, MSc

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Economics and Finance

QUT Business School

Queensland University of Technology

2016

Keywords

- Biased technical change
- Economic growth
- Education expenditure
- Health expenditure
- Human capital
- Inequality
- Optimal policy
- Political economy
- Stability analysis
- Steady state analysis
- Variable elasticity of substitution
- Wealth distribution
- Welfare

Abstract

This thesis consists of three essays that examine issues relating to human capital, economic growth, and inequality within the framework of dynamic macroeconomic models. A common innovation followed in all of these essays is the consideration of alternative approaches to the modelling of technology and preferences not typically observed in the extant literature, with a view to gaining insights into the diversity of growth and inequality patterns across countries.

The first study explores the properties of the variable elasticity of substitution production function, which is an appealing alternative to standard production functions considered in the literature, given that it is more general and nests several technological structures as special cases. It therefore has the potential to yield further insights on economic growth relative to existing studies. The dynamics of growth created by the variable elasticity of substitution production function is examined by incorporating it into an overlapping generations model à la Diamond (1965). We show that a certain technology parameter in the variable elasticity of substitution production function affects a number of features of the economy, such as the labour share in output, factor rewards, and capital intensity in production through the technical change it creates, and it may therefore provide a potential explanation for some of the changes in these features that have been observed in the recent growth literature. Growth dynamics are also determined primarily by this parameter. Depending on its value, the economy could reach a unique and stable steady state, fall into a poverty trap, or achieve an upper bound of capital stock per capita, as stipulated by the parameter restrictions associated with this particular form of production function. Hence, under the assumption of a variable elasticity of substitution production function, convergence towards a unique and stable steady state is not inevitable. The capital

biased technical change created by a higher value of this parameter is also shown to improve productivity in steady state, but causes an exacerbation of intergenerational inequality.

In the second and third essays, the focus shifts to human capital. The second essay examines an important dimension of human capital, namely education, and explores its implications for economic growth. In this study, we develop an overlapping generations model to explore how the degree of aggregate substitutability between public and private education expenditures, as measured by the elasticity of substitution between these two types of expenditures, impacts long run macroeconomic outcomes. The use of a variable elasticity of substitution "education production function", in which public and private education expenditures are the inputs, shows that greater aggregate substitutability between these two types of expenditure yields higher stocks of human and physical capital in the long run. Furthermore, the optimal tax rate is increasing in aggregate substitutability. Stability analysis reveals that the economy could reach a stable steady state characterised by either oscillatory or monotonic convergence, or display oscillatory divergence followed by the emergence of permanent endogenous cycles. The results of this study suggest that policies aimed at improving aggregate substitutability between public and private education expenditures could contribute towards improved long run macroeconomic outcomes.

The third study considers another dimension of human capital, namely health, and explores the implications arising from expenditures on health within an overlapping generations model with heterogeneous agents. In this model, a person's probability of survival into old age is determined by a variable elasticity of substitution health production function in which public and private expenditures are the inputs. Outcomes of this analysis are similar in spirit to the previous essay. Analytical and numerical results reveal that higher aggregate substitutability, as measured by the elasticity of substitution between private and public expenditures, leads to higher long run wealth levels and lower inequality. Specifically, in the political equilibrium, higher aggregate substitutability is associated with higher public expenditure on health. This study also concludes that policies aimed at increasing the degree of substitutability between private and public health expenditures could yield better long run outcomes.

Table of Contents

List of Figures

List of Tables

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

QUT Verified Signature

Signature:

Date: 28/07/2016

Acknowledgements

Many people have helped me in numerous ways during my PhD journey. First and foremost, I would like to extend my gratitude to my principal supervisor Dr Radhika Lahiri for being my guiding light. I thank her for being so generous with her time and her tireless mentoring. I am truly blessed to have had the opportunity to work with such an amazing person. I would also like to most gratefully acknowledge the guidance, feedback, and support provided by my associate supervisor Dr Oscar Pavlov throughout this journey. My entire PhD candidature has been a positive and fulfilling experience thanks to my extraordinary supervisory team.

I am grateful to QUT for providing me with a scholarship to undertake my PhD. If not for their patronage, I would never have been able to achieve my lifelong dream of completing my doctoral studies at a world class university.

I would also like to thank my friends and colleagues at QUT for their support and encouragement. I am especially grateful to Eucabeth, Thames, Sam, Suzanne, Zainab, Bushra, Uttam and others for their support and friendship. I am also grateful to all of the unit co-ordinators that I tutored for during the past six semesters, as well as the other staff members and research students at QUT who helped me in various ways and made my time at QUT a memorable phase of my life. I am also especially grateful for the valuable comments and feedback provided by Professor Michael Kidd, who chaired my confirmation panel, and as well as Professor Janice How and Dr Jay Sarkar, who were in my final seminar panel.

I would like to thank professional editor, Kylie Morris, who provided copyediting and proofreading services, according to the guidelines laid out in the university-endorsed guidelines and the Australian Standards for editing research theses.

Juggling family, part-time work, and a thesis is impossible without the support of an understanding spouse. I am grateful to my husband Nishada for foregoing many things over the past few years to support me to achieve my goal. I am blessed to have somebody like him to push me to face challenges and have faith in myself. Of course, our daughter Dinethra deserves special mention too. She has been the ray of sunlight who gave me the strength to go forward and appreciate that I did not need to look far to find happiness.

I am forever indebted to my parents for all they have done for me. Thank you for going beyond your means to educate me well, and thank you for making me the person I am today. If I could wish for one thing, it would be that my father would have lived long enough to see me complete my PhD. He inspired me to make the pursuit of knowledge my primary occupation and love learning for its own sake. I dedicate this thesis to him.

Chapter 1: Introduction

This thesis is an exploration of some of the issues relating to economic growth and inequality. Broadly, this thesis examines how technology, mechanisms for the accumulation of human capital, the political choices of citizens, and the initial distribution of wealth can affect an economy's growth patterns, inequality, and other long run economic outcomes. The motivational relevance for this work stems from the importance of these issues in macroeconomics. Although there is a large body of literature, both empirical and theoretical, around these issues, the underlying causes of poverty and inequality, and mechanisms of growth remain unresolved. This is witnessed through the diversity in growth and development experiences of world economies, the extent of which is captured by Prichett (2000) as follows:

> "While some countries have steady growth (hills and steep hills), others have rapid growth followed by stagnation (plateaus), rapid growth followed by decline (mountains) or even catastrophic falls (cliffs), continuous stagnation (plains), or steady decline (valleys)." (p. 221)

Given such diversity of growth experiences, a vast body of work exploring the factors that influence the process of economic growth and inequality has emerged over the past five decades or so. Much of this literature has been clustered around a number of prominent themes. For instance, the role and quality of institutions is an issue explored by, among others, Przeworski and Limongi (1993), Rodrik, Subramanian, and Trebbi (2004) and Acemoglu, Johnson, and Robinson (2005); while the role of geography has been the subject of studies such as Sachs (2001) and Olsson and Hibbs (2005). Authors including Glomm and Ravikumar (1992), Tamura (2006), and Galor (2011), on the other hand, consider human capital to be a primary determinant of economic growth and inequality. Technology has been the focus of studies such as those by McIntosh (1986), Levy (1981), and Acemoglu (2002), while studies by Yuhn (1991), Klump and De La Grandville (2000) and Palivos and Karagiannis (2010), among others, have examined the manner in which the elasticity of factor substitution affects growth patterns. Furthermore, the manner in which the financial sector influences the process of economic development has been the subject of studies such as those by Levine and Zervos (1998) and Roubini and Sala-i-Martin (1992).

While the list given above is not exhaustive, it demonstrates the fact that economic growth and inequality remain "open-ended" questions in the study of economics. Of the many determinants of growth and inequality, there has been considerable interest in the recent literature on technology, institutions, and human capital. This thesis considers some unexplored aspects of these dimensions. It consists of three studies that present theoretical perspectives on these matters through the development of approriate macroeconomic models. At the heart of the models presented in this thesis is a certain functional form called the variable elasticity of substitution (VES) production function developed by Sato and Hoffman (1968) and Revankar (1971). A version of this production function presented in Karagiannis, Palivos, and Papageorgiou (2005) is employed throughout the thesis. As it is a key building block in all of the models used in this thesis, the VES production function is essentially the common theme that binds these three studies together.

The distinctive feature of the VES production function is that the elasticity of factor substitution is linear related to the capital stock per worker. Depending on the value of a certain parameter, which could take a range of positive and negative values, the elasticity of factor substitution could depend positively or negatively on the per capita capital stock. This results in the elasticity of factor substitution taking different values at different points on a particular isoquant. Given that a higher capital stock per worker is closely associated with economic growth, the elasticity of factor substitution is *endogenous* in the VES form. Such endogeneity of the elasicity of factor substitution has been suggested by, among others, Miyagiwa and Papageorgiou (2007) and Yip and Xue (2013). As will be shown, this endogeneity enables the VES production function to be adapted to explore the macroeconomic outcomes associated with the public-private mix of merit goods such as health care and education.

We now provide an overview of the three studies that constitute this thesis. The first essay introduces the VES production function, discusses its properties, and looks at the growth dynamics it creates in the context of an overlapping generations model á la Diamond (1965). The main motivation for the study is the empirical evidence in studies such as Arrow, Chenery, Minhas, and Solow (1961), Bairam (1989), and Duffy and Papageorgiou (2000), which point to considerable variations in the value of the elasticity of factor substitution between countries and industries. In view of such evidence, as suggested by Miyagiwa and Papageorgiou (2007), the elasticity of factor substitution may be an *endogenous* variable that may be linked to the level of economic development of a country. The constant elasticity of substitution (CES) production function and its special case, the Cobb Douglas form, which are usually used in most theoretical and empirical studies on growth and inequality, cannot take such endogeneity into account, as the elasticity of factor substitution remains static over time and input combinations in these functional forms. Hence, there is a need to consider more flexible, alternative forms of production function such as the VES form. In the VES form, the elasticity of factor substitution depends on the capital to labour ratio used in production, and its value therefore varies along different points of an isoquant. As the capital to labour ratio is directly linked to an economy's rate of growth, the VES form creates a link between the elasticity of substitution and economic growth.

Further motivation for the study emerges from some of the recent observations relating to economic growth, such as the global decline in the labour share and the increase in the returns to capital, for which the standard Solow-Swan growth model under the assumption of a CES production function cannot provide adequate explanations. The central parameter in the VES production function referred to earlier, which is labelled b throughout the thesis following Karagiannis et al. (2005), provides a potential explanation for such recent observations relating to the growth experiences of countries.

In particular, in the spirit of Acemoglu (2002), we show that this parameter is a source of *biased technical change*. Specifically, a higher value of b causes the marginal product of capital to rise, while the marginal product of labour falls concurrently. These movements result in greater bias in production, which in turn influences the elasticity of factor substitution. In the VES form, a positive value of b yields an elasticity of substitution above 1, and a negative vaue of b being associated with an elasticity of substitution between 0 and 1. The value of b also influences relative factor shares, with higher values of this parameter leading to a higher share of capital in total output. Given that this parameter is a key determinant of a number of characteristics of the economy it can be interpreted as an important structural parameter. A higher value of b essentially represents a technology that enables greater factor substitution, and is hence determined by a number of factors, such as the power of trade unions, the attitudes of the government towards investments in different factor inputs, as well as developments in the financial sector that determine investments in new technology.

Upon exploring the properties of the VES production function, we proceed to apply it to an otherwise standard Diamond overlapping generations model, where agents live for two periods, working during the first and spending the second in retirement. Agents save in the first period in order to finance their old age consumption, and these savings make up the capital stock of the next period. The key results of this exercise are that a positive value of the parameter b is always associated with a unique and stable steady state analogous to that associated with the standard Solow-Swan model. When the value of b is negative, however, there are three possible long run outcomes: the economy could achieve a stable steady state, fall into a poverty trap, or reach an upper bound of per capita capital stock. The VES form cannot produce unbounded growth given the parameter restrictions that are imposed to ensure that the elasticity of factor substitution and the marginal product of capital are positive.

We show that a higher value of the parameter b enables the economy to produce a higher level of output per capita at steady state with a lower per capita capital stock. Hence, a higher value of this parameter creates an efficiency effect. However, in the presence of competitive factor markets where the reward for each factor input is equal to its marginal product, a higher value of this parameter, by causing the marginal product of capital to rise and marginal product of labour to fall, can lead to an exacerbation of intergenerational inequality, as the old generation owns the capital and the young generation supplies the labour needed for production in a Diamond overlapping generations model.

The second and third studies of the thesis are *applications* of the VES production function in the contexts of education and health respectively. In these studies we examine how the degree of substitutability between public and private expenditures on education and health can impact on long run macroeconomic outcomes. The functional form and properties of the VES production function enables easy adaption to explore these issues. In the second and the third studies, the VES is adapted to define health and education "production functions" in which public and private expenditures form the inputs. The use of the VES production function in the extant literature has been restricted primarily to "standard" contexts where two inputs are employed to produce a consumer good. To our knowledge, this is the first time the VES form has been used to study the accumulation of human capital and the resulting dynamics.

Much of the extant literature dealing with public-private mix of expenditures in markets like health care and education assumes either rigid complementarity or perfect substitutability between these two inputs. The key motivation for the second and third studies is the fact that, in practice, the degree of substitutability between these two inputs does not admit to either of these extremes. The extent to which they are substitutable or complementary to each other would depend on a range of cultural and institutional factors. Furthermore, the degree to which an individual perceives these inputs to be substituable or complementary would also depend on several other factors such as the person's income, location, and access to information. Only a handful of extant studies, such as Bearse, Glomm, and Patterson (2005), Bhattacharya and Qiao (2007), Osang and Sarkar (2008), Lahiri and Richardson (2009), and Li, Moslehi, and Yew (2012) have considered the impacts on growth, inequality, and political economy outcomes arising from different levels of substitutability between public and private expenditures in the contexts of health and education.

It is this strand of literature that is of direct relevance to the second and third studies of this thesis. The use of the VES form enables us to consider a range of intermediate values of the elasticity of substitution between public and private expenditures in these markets. A further advantage of using the VES form is that it improves the analytical tractability of our results.

In the second study, we consider a three period overlapping generations model where the agent spends her childhood acquiring human capital, her adult age working and her old age in retirement. In adult age she gives birth to a single offspring. The offspring's education is funded by the government, which imposes a tax on the adult agents of the economy in order to finance the public education system, as well as by the parents. We define an education production function of the VES type in which these public and private expenditures form the inputs.

In this context, the parameter b captures a number of institutional and cultural factors that influence the degree of substitutability between public and private education expenditures. As elaborated in Chapter 4, these factors involve the cultural emphasis on educational attainment, which might motivate parents to supplement public education with private expenditures in order to allow their children to achieve better academically, as well as the actual quality of the public education system and the various regulations relating to private sector involvement in the education industry. Given that it captures a number of such considerations, we refer to the parameter b as the *aggregate substitutability parameter*. This parameter can also be viewed as a representation of the relative uniformity between public and private education expenditures in comparison to private education, with a higher value of b being associated with a greater degree of uniformity between public and private education expenditures. A higher value of this parameter therefore indicates greater comparability of these two inputs in relation to coverage, access, and quality.

The agent's adult age human capital is assumed to depend on the education she acquired in childhood, as well as her parent's human capital, according to a Cobb-Douglas specification with increasing returns to scale. Production in this economy also follows a Cobb-Douglas form, with human and physical capital forming the inputs.

We show that the steady state stocks of human and physical capital are rising in the value of b, and that a higher value of this parameter enables the economy to converge to the steady state faster. Hence, greater substitutability between public and private education expenditures is desirable for the economy. It is also interesting to note that the ratio of physical to human capital rises in the aggregate substitutability parameter.

We carry out stability analysis to gain insights into the different transition dynamics that could occur. This analysis reveals that the economy could display monotonic or oscillatory convergence towards the steady state or display oscillatory divergence away from the steady state followed by the emergence of permanent endogenous cycles.

The third essay uses the VES form to extend the idea of substitutability between public and private expenditures to the context of health. It examines how the degree of substitutability between public and private health expenditures influences wealth dynamics and political economy outcomes. In this study, we consider a two period model where agents have heterogeneous initial endownments of wealth. In both youth and old age, they earn exogenous wages. In youth, the agent gives birth to a single offspring and the youthful agent also inherits a bequest from her parent. The government charges a proportional tax on the bequests received by young agents. A certain proportion of the total tax revenue is used to finance public health care, while the remainder is redistributed equally among agents in the form of a lump-sum transfer. Young agents use their net income to fund private health expenditures and savings. Public health expenditure and the spending on health undertaken by agents in youth determine the stock of health capital an individual possesses at the end of her youth according to a VES form, and this stock of health capital determines her probability of survival into old age. This survival probability function is a non-convex function of health capital. If the agent safely survives through old age, she allocates her income from work and her savings that accumulate an exogenous interest on consumption and bequests for her offspring. If the agent does not survive through old age safely, her offspring inherits the gross value of her parents' savings as an unintended bequest. We assume that savings or unintended bequests are always less than intended bequests for all agents in the economy.

When applied in this context, the parameter *b* captures all of the aggregate-level factors that affect the ease with which an individual can substitute between private and public health care expenditures. Along a similar vein to the second essay, it is an indicator of the comparability of the public and private components of a health care system, with a higher value of *b* being associated with a greater degree of similarity between public and private health expenditures along dimensions such as quality and access. The degree of substituability at the aggregate level naturally improves an individual's ability to substitute between public and private health expenditures. However, using the VES form allows the individual's ability to substitute between public and private health expenditures to also depend on the actual ratio of public to private health expenditures she undertakes, which is directly dependent upon her wealth. The heterogenous agents framework considered enables us to make a distinction between both aggregate and *individual* substitutability of health expenditures. Hence, we make a novel contribution to the literature by directly incorporating micro-level considerations relating to heterogeneity in substitutability into our model, and examine how these aggregate and micro-level factors can affect wealth inequality and macroeconomic performance in the long run.

The main results of this study are as follows: higher aggregate substitutability between public and private health expenditures is associated with higher average wealth levels and lower inequality. This outcome occurs due to the political economy determination of the proportion of tax revenue that agents desire to be allocated towards public health care. In the presence of greater aggregate substitutability, agents vote for a higher proportion of tax revenues to be allocated towards public health care, which enables agents to achieve a particular survival probability by undertaking less private health expenditure and leave more bequests (intended or unintended) for their offspring. Hence, higher public health care expenditure acts as a mechanism for the redistribution of wealth in the economy.

Another interesting dimension explored in this study is how the political economy outcomes compare with the alternative determination of public health spending by a benevolent social planner who wishes to maximise the welfare of agents. To this end, the political economy outcome is compared to the social welfare maximising outcomes based on the utilitarian and Rawlsian paradigms. It can be seen that for a large range of parameter values, the political economy, Rawlsian, and utilitarian outcomes coincide in the long run. In the instances they do not, the political economy outcome converges to either the Rawlsian or utilitarian outcome, suggesting that the political economy outcome is generally a socially desirable one. This comparison is another interesting contribution of the study, through which the welfare implications of the political economy outcomes along the economy's transition path are ascertained by comparing them against two opposing perspectives on welfare.

The diverse explorations carried out in the three essays provide interesting insights into the underlying factors that determine the manner in which growth, inequality, and political economy outcomes evolve in an economy over time. This thesis makes several novel contributions to the extant macroeconomics literature. The explorations of the first study aim to contribute towards the recent growth literature, especially the strands of literature that deal with the manner in which the elasticity of factor substitution affects growth outcomes and the direction of technical change. In particular, the VES form provides a means of capturing the variations in factor substitutability that has been documented in a large number of empirical studies. Furthermore, this thesis reinforces the beneficial impacts of a higher elasticity of substitution demonstrated in the extant literature, albeit through a different channel: through its influence on the direction of technical change in the economy the parameter b has an impact on the elasticity of factor substitution, and in turn the dynamics of growth.

The second and the third studies contribute towards the literature on human capital accumulation, economic growth, and inequality. In the second study, the VES education production function is used to explore the dynamics of physical and human capital accumulation in the economy. One of the few studies that have explicitly looked at the issue of substitutability in the context of education is Bearse et al. (2005). However, in contrast to their study, which looks at the political economy outcomes, our focus on the dynamics of human and physical capital accumulation, enables us to make an interesting addition to the growth literature. In the third study, the use of the VES health production function enables a distinction between individual and aggregate substitutability in this context, which is an aspect extant studies exploring this issue, such as Bhattacharya and Qiao (2007), Lahiri and Richardson (2009) and Li et al. (2012), have not taken into consideration. Furthermore, the comparison of the political economy outcome for the proportion of tax revenue to be allocated towards providing public healthcare with the value selected by a social planner under both the Rawlsian and the utilitarian paradigms makes an important contribution to the implications on welfare that emerge from the issue of substitutability between public and private expenditures in the context of health.

This thesis also carries several policy implications. In general, the thesis highlights the beneficial impacts associated with a higher degree of substitutability between factor inputs, as well as between public and private inputs in health care and education. In all of the contexts considered in this thesis, government policies directed towards raising the value of the aggregate substitutability parameter b can yield better long run outcomes. With regard to the standard form of the VES production function considered in the first essay, as noted previously, the parameter b represents a technology that, through the capital bias it creates, enables greater factor substitution. Hence, in this context, certain government policies could influence the value of b. For instance, certain political parties may promote capital bias by encouraging investments that raise the relative price of capital. Furthermore, as a higher value of b is associated with a lower share of capital in output, the power of trade unions could also be a key determinant of its value, as trade unions typically attempt to maintain the share of labour at a high level. Hence, reforms aimed at reducing the power of trade unions could also contribute towards raising the value of the parameter b. Nevertheless, it should be noted that a while a higher value of the central parameter enables the economy to reach a higher value of output per capita at steady state, this occurs at the expense of higher intergenerational inequality. Hence, a government that is especially concerned about the worsening inequality created by a higher value of this parameter may settle for a set of policies that keep this parameter at a comparatively lower level.

In the contexts of education and health, as mentioned earlier, b captures a range of cultural and institutional factors that affect peoples' perceptions regarding the extent to which public and private expenditures are substitutable for each other. Hence, policies directed towards these factors could help raise the value of the parameter b. In the case of education, these policies could include derregulation of the industry to enable private providers to provide high quality education services in countries where there is considerable competition for publicly provided education. Nevertheless, in contexts where a private education is perceived to be of a superior quality, improving the quality of public education could improve the comparability of these two types of inputs. In the case of health, encouraging private participation in heavily regulated markets and improving the quality of public healthcare, especially in developing countries where publicly provided health services are of a considerably lower quality relative to their private counterparts, could also help raise the value of the aggregate substitutability parameter.

However, with both health and education, it is important to bear in mind that many of the factors, especially cultural ones, that affect the aggregate substitutability parameter are typically sticky, and the influence government policy could exert upon them could be limited. Through further empirical investigation of the key outcomes of our work, it may be possible to develop more specific and comprehensive policy recommendations that could potentially play an important role in improving long run economic outcomes in these contexts.

The rest of this thesis is organised as follows: Chapter 2 reviews some literature related to the studies included in the thesis. Chapters 3, 4 and 5 contain the first, second and third essays respectively, and the concluding remarks are provided in Chapter 6. Proofs and derivations are relegated to the Appendices.

Chapter 2: Related Literature

2.1 INTRODUCTION

This chapter focuses on a number of themes of relevance to the studies that constitute this thesis. Given that this thesis consists of three independent essays, the issues that are of immediate relevance are reviewed in each study separately. The aim of this chapter is to provide an overarching background that connects themes appearing in the individual essays. As the first essay explores the VES production function in the context of a model of economic growth, a survey of the recent literature that looks at patterns of economic growth, overall growth trends, as well as certain regional and country level growth experiences is an apt starting point for this chapter. This brief account of growth patterns comprises Section 2.2. Given that economic growth was regarded as an "open-ended question" in the Introduction, exploring the causes of economic growth is obviously an important consideration. Hence, a number of different causes of economic growth are examined in Section 2.3. As the second and third studies of this thesis examine issues relating to growth and inequality in the contexts of education and health respectively, we define human capital and discuss its different components in Section 2.4. In particular, attention is directed towards the idea that, like education, health is also an important type of human capital. Section 2.5 investigates some of the intricate linkages between human capital, inequality, and economic growth. Section 2.6 concludes the chapter.

2.2 GROWTH PATTERNS AROUND THE WORLD

The Industrial Revolution is often considered to be one of the most important landmarks in modern economic history. While the period beyond this point has been characterised by notable improvements in average incomes and living standards, the fact that Europe and its offshoots grew richer while many countries lagged behind caused the emergence of large global income inequalities (Baldwin, Martin, & Ottaviano, 2001; C. I. Jones, 2001). European countries enjoyed an even greater boost to their growth rates after World War II, and this era also marked the emergence of Japan as a world economic superpower, as it rapidly caught up with the original leaders in the world economy (Abramovitz, 1990; De la Fuente, 1997). In the 1980s, Hong Kong, Singapore, Taiwan, and South Korea, better known as the Asian Tigers, embarked on a speedy trajectory of growth and joined the most economically powerful nations in the world.

In recent years, with its remarkable growth rate averaging around 10% per annum, China has cemented its position as a powerful entrant into the catch-up game (Crafts, 2004). India has also been a powerful force in the world economy; until the 1980s India's trend rate of growth was around 3% per annum, but by 2007, this rate had increased to approximately 9.7% (Bollard, Klenow, & Sharma, 2013). If they continue their impressive growth trajectory, China will become the second largest economy in the world by 2016, and India is projected to become the world's third largest economy by 2035 (Kaplinsky & Messner, 2008).

Alongside a number of older and more recent success stories, there are many other countries that have experienced modest, if not negative, growth. This is illustrated by Temple (1999), who uses data for the 1960-1990 period to distinguish between "growth miracles," which are countries that enjoyed continuous improvements in per capita income during the time period considered and "growth disasters," which are countries that experienced a decline in per capita income. According to his classification, South Korea, Japan, Botswana, Hong Kong, Taiwan, Singapore, Malta, Cyprus, Seychelles, and Lesotho are growth miracles; while Ghana, Venezuela, Mozambique, Mauritania, Nicaragua, Zambia, Mali, Madagascar, Chad, and Guyana are growth disasters.[1] With regard to such a classification of countries, C. I. Jones (1997) notes that in recent decades, growth miracles have been more common than growth disasters, and therefore predicts that a marked improvement in the income levels of people in many countries would occur in the long run.

A more elaborate and extensive classification of growth patterns covering a large number of countries is provided by Pritchett (2000). He distinguishes between growth paths that resemble hills, steep hills, plateaus, cliffs, mountains, plains, and valleys. Hills and steep hills represent steady growth over time, plateaus capture a scenario where a country experiences rapid growth followed by stagnation, a cliff is a catastrophic fall in income levels, a mountain is rapid growth and subsequent decline, plains capture stagnation of growth, "accelerators" are countries that experience little or no growth at first but then display accelerated growth, while valleys capture a steady decline in growth over time. Pritchett's classification of countries based on this analogy is provided in Table 2.1.

[1] As this classification is based on an older dataset, it does not apply well in the recent context.

Table 2.1: Pritchett's classification of countries

Group	Countries
Steep hills	Japan, Thailand, Cyprus, Ireland, Taiwan, South Korea, Botswana, Malaysia, Singapore, Malta, and Hong Kong
Hills	Australia, Austria, China, Myanmar, Bangladesh, Israel, Tunisia, Barbados, Tanzania, Belgium, Canada, Philippines Pakistan Turkey Colombia, Costa Rica, Denmark, Finland, Mexico, France, Germany, Greece, Italy, Portugal, Spain, Switzerland, United States
Mountains	United Kingdom Namibia, Papua New Guinea, Algeria, Egypt, Argentina, Cameroon, Congo, Iran, Iraq, Bolivia, Ecuador, Cote d'Ivoire, Jordan, Saudi Arabia, Guyana, Gabon, Liberia, Syria, Honduras, Mozambique, Jamaica, Niger, Nigeria, Nicaragua, Sierra Leone, South Africa, Panama, Togo, Paraguay, Peru, Zaire, Zambia, Suriname, Trinidad and Tobago
Plateaus	Iceland, Morocco Brazil, Ethiopia, Netherlands, Dominican Republic, Gambia, Guinea Bissau, New Zealand, Sweden, Bissau, Kenya, El Salvador, Lesotho, Malawi, Guatemala, Swaziland
Accelerators	Indonesia, India, Sri Lanka, Chile, Uruguay Ghana, Mauritius
Plains	Nepal Haiti, Republica Angola, Burundi, Benin, Central African Republic, Venezuela, Guinea, Burkina Faso, Madagascar, Mali, Mauritania, Rwanda, Senegal, Somalia, Uganda, Zimbabwe

Adapted from Pritchett (2000)

The fact that Pritchett's (2000) analysis does not classify any countries in the cliffs and valleys categories projects some optimism in relation to growth prospects of the countries considered. Nevertheless, this classification is based on older data, and the growth experiences of certain countries have changed since, thereby prompting the need to interpret it with caution. For instance, as a country that was severely affected by the global financial crisis, Greece's growth rates have been consistently negative since 2008, although in 2014 its growth rate was 0.77%, which projects a slight recovery (World Bank, 2015). Hence, in recent years, Greece's economic growth pattern has borne a closer resemblance to a mountain than a hill. Another example is

Syria; the World Bank has provided no growth data for Syria from 2008 onwards, but since the civil conflict erupted, Syria's growth pattern may well have changed from a mountain to a cliff. Such experiences demonstrate that the growth patterns of an individual country may display considerable variations over time. Such changes in growth paths in a particular country over time have been studied, for instance, by Jerzmanowski (2006), who builds a Markov-chain model drawing on Pritchett's work to show that countries may switch between growth regimes over time, with the probability of such transitions being determined by the quality of institutions. His model comprises of four growth regimes: stable growth, miracle growth, stagnation, and crisis. Low institutional quality diminishes a county's ability to sustain stable growth. Adding to this sub-strand of literature, Kerekes (2012) interprets economic growth as a "sequence of transitions between different growth regimes that countries visit with different frequencies" (p. 167) and builds a Markov-chain switching model using data from 82 countries covering the period 1962 to 2003 to demonstrate that countries switch between regimes of stable growth, miracle growth, stagnation, and crisis. She attributes a successful growth record to initial conditions prevailing in a particular country, policies undertaken by authorities, and institutional quality.

When reflecting on these diverse growth patterns that can be observed between and within countries over time, an important question one should ultimately ask is whether the developing countries can ever attain the levels of economic growth and living standards of their developed counterparts. To provide a prelude to this matter, the average annual growth rates for 27 OECD countries, the BRICs and a large number of less developed countries for the period 1991-2014 are presented in Table 2.2. It is interesting to note that only a few OECD countries, namely Ireland, Chile, South Korea, and Israel, have been able to maintain average growth rates above 4% during this period. Compared to this, China's phenomenal average growth rate of 10.09% and India's growth rate of 6.56% lends support to the notion of catch-up or "convergence". This is because, if they continue to maintain such growth rates, they could, in the long run, reach levels of economic development comparable to those of industrialised countries. As Bollard et al. (2013) suggest, such a catch-up process could have implications for development and welfare globally; given the large populations and the high prevalence of poverty in India and China, the high economic growth rates in these two countries could help transform the lives of a marked proportion of the world's poor.[2]

[2] Given that China and India contain 20% and 17% of the world's population respectively, their growth has a marked impact on the global economy as well. Hence, studying their growth experiences requires one to deviate from the "small economy" assumption usually made in growth models (Kaplinsky & Messner, 2008)

Table 2.2: Annual average growth rate of GDP between 1991 and 2014

Country	Average annual growth rate of GDP	Country	Average annual growth rate of GDP
OECD countries		*LDCs*	
Australia	3.11	Burundi	1.34
Austria	1.91	Benin	4.34
Belgium	1.70	Burkina Faso	5.83
Canada	2.39	Central African Republic	0.40
Chile	5.03	Comoros	2.21
Czech Republic	1.70	Eritrea	3.73
Germany	1.45	Ethiopia	6.57
Denmark	1.47	Guinea	3.13
Spain	2.00	Gambia, The	3.26
Finland	1.71	Guinea-Bissau	1.97
France	1.52	Liberia	9.08
United Kingdom	2.01	Madagascar	2.30
Greece	1.03	Mali	4.69
Ireland	4.35	Mozambique	6.73
Iceland	2.65	Malawi	4.15
Israel	5.05	Niger	3.77
Italy	0.65	Nepal	4.45
Japan	0.92	Rwanda	5.79
Korea, Rep.	5.12	Sierra Leone	3.84
Luxembourg	3.58	Chad	6.58
Netherlands	1.91	Togo	2.88
Norway	2.47	Tanzania	5.18
New Zealand	2.74	Uganda	6.72
Poland	3.70	Congo, Dem. Rep.	0.98
Portugal	1.25	Zimbabwe	0.24
Sweden	2.05		
United States	2.48		
BRICs			
Brazil	2.98		
China	10.09		
India	6.56		
Russian Federation	0.95		
South Africa	2.60		

Source: World Bank data

It is also encouraging to observe that certain less developed countries, such as Ethiopia, Liberia, Chad, Mozambique, and Burkina Faso, have been able to maintain average growth rates above 5% during the period considered in Table 2.2, which points to an improvement in the living standards of the citizens of these countries during the period. Evidence suggesting that such economic growth can be pro-poor is provided by Ravallion (2001), who uses data from 47 developing countries to show that a 1% increase in mean income leads to a 2.5% reduction in the proportion of people living on less than $1 a day.

However, many countries, such as Zimbabwe, the Central African Republic, the Democratic Republic of Congo, and Russia have displayed sluggish growth during this time, suggesting that income disparities between these countries and the developed world are only widening.

At the heart of these diverse growth experiences lies the convergence debate—the question of whether poorer countries are catching up with their rich counterparts, which, according to Romer (1994) is the issue that has been most extensively explored by growth economists in recent times. Over the past few decades, three major perspectives on convergence have emerged in the empirical and theoretical growth literature: absolute convergence, conditional convergence, and club convergence.

The idea of absolute convergence propagates the view that, regardless of initial conditions, all economies will converge to a common level of per capita income in the long run. For instance, Barro and Sala-i-Martin (1992) evidence an absolute convergence of incomes among the states in the US, as well as in a sample of 48 countries. Much of the earlier empirical literature estimated that developing economies were converging with their developed counterparts at a rate of around 2% per annum (Quah, 1996), suggestive of an absolute convergence of incomes on a global scale.[3]

[3] However, Quah (1996) suggests that this may be a technical artefact emerging from the possible presence of unit roots in the time series regressions considered in these studies.

The conditional convergence hypothesis is based on the standard model of economic growth developed by Solow (1956) and Swan (1956), which predicts that, under certain standard assumptions, an economy always monotonically converges to a unique and stable steady state.[4] This steady state value is conditional in that it is determined by the rates at which population growth, technological progress, savings, and depreciation of the capital stock occur in a particular country. Once the economy reaches steady state, output per effective worker remains constant, implying that the output per worker increases at the rate at which exogenous technological growth occurs. Empirical evidence to support the conditional convergence hypothesis has been provided by, among others, Barro and Sala-i-Martin (1992), Mankiw, Romer, and Weil (1992), and Islam (1995).

The third perspective is the idea of club convergence, which proposes that countries with similar initial incomes and structural features will cluster around the same per capita income in the long run, leading to polarisation of cross country incomes (Galor, 1996). A large number of studies have suggested that the highly industrialised countries (i.e., the OECD member countries) experienced a convergence of productivity and incomes during the period between 1870-1990, while less developed countries fell far behind their developed counterparts, with the ratio of average incomes of the richest and poorest nations of the world rising nearly five-fold during the period (Pritchett, 1997). This observation demonstrates a clear segregation of countries into a developed group and a developing one, and thereby provides support for the club convergence hypothesis. This view is further expounded by Quah (1996), who provides empirical evidence that shows a gradual polarisation of cross

[4] These standard assumptions are continuity, differentiability, diminishing marginal products, constant returns to scale, and that the Inada conditions are satisfied.

country incomes that leads to two distinct clubs: one consisting of rich countries and the other consisting of poor countries.

The presence of empirical evidence to support all three perspectives on convergence means that much debate remains around this issue among macroeconomists. However, summing up the current state of affairs, Piketty and Goldhammer state that:

> "A global convergence process in which emerging countries are catching up with developed countries seems well under way today, even though substantial inequalities between rich and poor countries remain." (2014, p. 72)

The reduction in cross-country inequalities resulting from the global convergence process is likely to transform the lives of millions of people worldwide. However, the substantial inequalities that still persist may mean that the fruits of economic growth are unlikely to ever be distributed equitably, both across and within countries. Hence, inequality is naturally intertwined with economic growth. The subject of inequality is revisited in greater depth in Section 2.5.

Nevertheless, if these variations and diversities in economic growth rates between and within countries are averaged out and a "very long run" perspective is adopted, the global growth in per capita income has actually been very modest. Picketty and Goldhammer's (2014) calculations of the average annual growth in per capita output over the past two millennia, which are summarised in Table 2.3, show that for the first 1700 years of modern human history, output remained stagnant. The best global economic performance occurred in the period 1913-2012, when average annual output grew at a rate of 1.6% on average. The data is supportive of the view that economic growth is a relatively recent phenomenon that came into existence as late as the 19th century. Despite these modest values however, it is important to appreciate that an annual growth rate of 1.6% can culminate into a very large increase in income per capita over a long period of time, thereby demonstrating the extent to which economic growth can make a positive contribution towards enriching the lives of people on a global scale.

Table 2.3: Average annual growth in per capita output 0-2012

Years	**Average annual growth rate of per capita output (%)**
0-1700	0
1700-1820	0.1
1820-1913	0.9
1913-2012	1.6

Adapted from Piketty and Goldhammer (2014, p.73)

2.3 ECONOMIC GROWTH: PROXIMATE AND FUNDAMENTAL CAUSES

In the Solow-Swan model, a country's growth rate is determined by the rate at which it accumulates human and physical capital and its technology progresses. This close relationship between economic growth and these variables posited by the Solow-Swan model is an empirical regularity that has been observed in many contexts. There are several perspectives on estimating the contributions of these three variables. For instance, growth accounting estimates the growth rate of output implied by growth rates in physical and human capital, and attributes the difference between the implied and actual rates of growth to total factor productivity (TFP), or the "Solow residual" as it is commonly known in the growth literature (Baier, Dwyer, & Tamura, 2006). TFP is a measure of how efficiently human and physical capital are used in production, and therefore captures the long run technical change in an economy. On the other hand, endogenous growth theory assumes that human capital is the key driver of economic growth, as a country's stock of human capital affects the growth process *through* the contributions it makes towards technological development (Benhabib & Spiegel, 1994).

Despite the major focus in the growth literature on technological progress, human, and physical capital, Acemoglu (2008) asserts that they are only *proximate causes* of economic growth, because investments in these factors undertaken by people and governments are governed by certain "deep" characteristics, which he refers to as the *fundamental causes* of economic growth. He identifies these fundamental causes of growth as: luck, institutions, geography and culture. One of the first proponents of some of these factors as fundamental determinants of growth were Nelson and Phelps (1966), who noted that the technology parameter in the Solow growth model encapsulates institutions, soil quality, and natural resource endowments, among other things. The next sections discuss these fundamental causes of growth and then examine a number of additional factors identified in the recent growth literature as closely related to growth.

2.3.1 The luck hypothesis

As discussed previously, the varied patterns of growth observed in practice are suggestive of the idea that, unlike in the Solow-Swan model, which predicts conditional convergence towards a unique and stable steady state, countries are actually confronted with multiple equilibria in practice. In the presence of multiple equilibria, small differences between countries with largely similar characteristics could yield very different long run growth paths.

To demonstrate this more elaborately, Figure 2.1 shows a growth trajectory with multiple equilibria. A country that starts with a per capita capital stock below k_1 falls into a poverty trap, while a country with an initial capital stock between k_1 and k_2 achieves a steady state capital stock of k_2. Finally, a country that has an initial capital stock above k_2 will be able to enjoy unbounded growth.

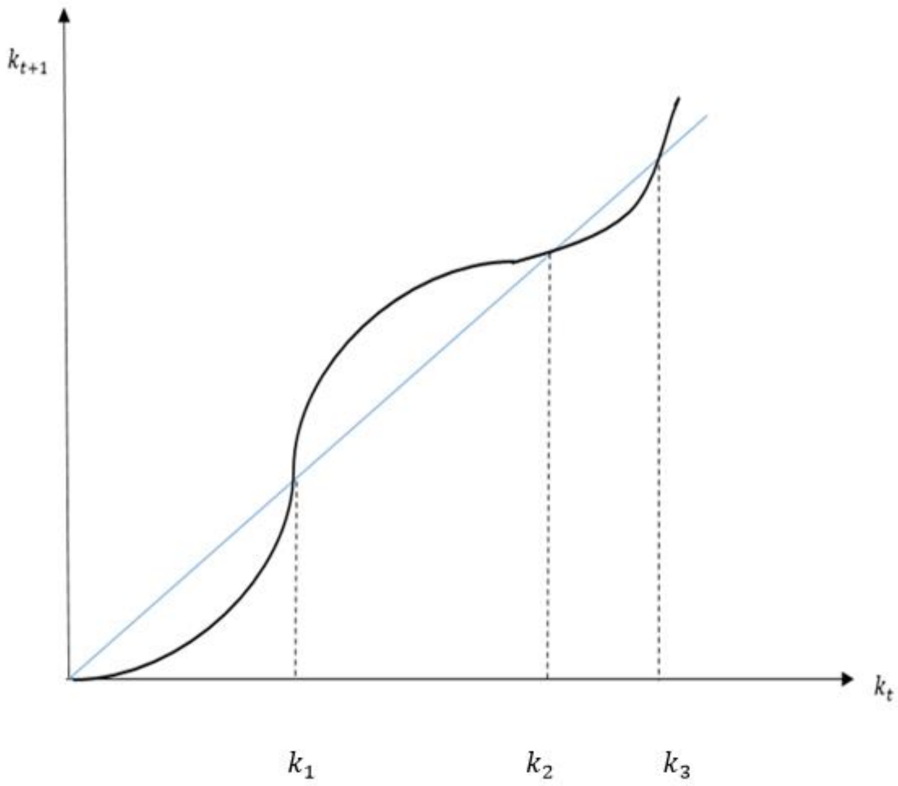


Figure 2.1 Economic growth with multiple equilibria

Adapted from de la Croix and Michel (2002, p. 33)

Acemoglu (2008) makes a distinction between models with multiple equilibria and those with multiple steady states with history dependence. In the former case, a change in the beliefs and behaviours of people can enable a nation to transition from a sub-optimal equilibrium path towards a Pareto efficient one, while in the latter instance, once a country embarks on a particular growth trajectory, digressing from it is impossible. In the case of multiple steady states with history dependence, the long run outcomes that transpire depend on the initial stock of capital (human or physical) that a country possesses. Hence, if the dynamics depict vastly different outcomes, the history dependence view suggests that the fortunes of a country depend entirely upon its initial stock of capital. Such a dependence on historical factors may account for the persistent differences in incomes between countries. On the other hand, growth miracles, such as Japan, South Korea, Singapore, and more recently, the impressive growth rates recorded by China and India, point towards the possibility of transition from one equilibrium to another. However, growth models with multiple equilibria shed no light on the factors that force a country to remain trapped in a sub-optimal equilibrium or enable it to escape from such an equilibrium. The other fundamental causes of growth could potentially provide explanations for such issues.

2.3.2 The geography hypothesis

The geography hypothesis proffers that geographical, climatic, and topological conditions can be a primary determinant of a country's growth experience. Soil quality and climate can have a direct impact upon agricultural productivity, and a country endowed with mineral resources may be able to use the revenue from the sale of these extracts to support industrial development (Acemoglu, 2008). For instance, Canada has a climate that helps to contain the spread of diseases and is conducive to agricultural production, is bordered with a prosperous trading partner, and is sea-navigable. These geographical characteristics may be a possible explanation for Canada's emergence as one of the richest countries in the world (Easterly & Levine, 2003).

Jeffrey Sachs, a pioneer of the geography hypothesis asserts that, "Perhaps the strongest empirical relationship in the wealth and poverty of nations is the one between ecological zones and per capita income. Economies in tropical ecological zones are nearly everywhere poor, while those in temperate ecological zones are generally rich." (Sachs, 2001, p. 1). In relation to this point, he notes that the only developed countries located in the tropics are Hong Kong and Singapore. Furthermore, climatic conditions affect the spread of diseases. For instance, there are many diseases like dengue, malaria, and yellow fever that are specific to tropical climates, and their prevalence can have an adverse impact on productivity.

The location of a country may also affect its growth potential. This is demonstrated by MacKellar, Wörgötter, and Wörz (2000), who reveal that, on average, the growth rate of a landlocked country is about 1.5 percentage points lower than that of a sea-navigable country. Landlocked countries are often also at a disadvantage in international trade, as their ability to transport the goods produced domestically to the rest of the world is heavily dependent upon the transport infrastructure of transit countries, as well as the maintenance of amicable international relations with these countries. The fact that landlocked countries engage in less international trade is highlighted by Limão and Venables (2001), who use World Bank data to show that in 1995, landlocked countries' share of imports in gross domestic product was around 11 percent, while it was 28 per cent for coastal economies. Furthermore, during the period 1965-1990, none of the top 15 exporter of non-primary products were landlocked nations. Nevertheless, the potential benefits that regional integration can provide to landlocked countries have been demonstrated by the economic success achieved in recent times by landlocked countries of Western and Central Europe such as Austria, the Czech Republic, Hungary, the Former Yugoslav Republic of Macedonia, Slovakia, and Switzerland (Gallup, Sachs, & Mellinger, 1999).

From a historical perspective, the presence of animals and plants suitable for domestication enabled societies to transition from the hunter-gatherer stage to agricultural societies, and empirical evidence suggests that such prehistoric biogeography is an initial condition that supports a country's modern economic development (Olsson & Hibbs, 2005). For instance, J. Diamond (1997) identifies nine regions in the world where domestication of flora and fauna occurred independently and three more where local food production was initiated by "founder" crops with their origins in other regions, while a number of other regions never adopted agricultural production until they were compelled to so in modern times. The earlier onset of agricultural production enabled certain regions to achieve greater economic development in modern times.

2.3.3 The culture hypothesis

The third fundamental cause of economic growth identified by Acemoglu (2008) is culture. In the economics literature, culture is often defined as "individual and collective beliefs, social norms, and various attributes of individuals' preferences that are somehow influenced by their environment, but typically slow moving." (Aghion & Howitt, 2009, p. 421).

That culture is a primary determinant of economic progress is a view more than a century old. For instance, in his iconic book, *The Protestant Ethic and the Spirit of Capitalism*, first published in 1904, Max Weber argued that the Protestant Reformation had a major influence on the spread of capitalism in Western Europe and the region's economic prosperity. This hypothesis has been empirically tested by a number of authors. For instance, that Protestant countries in the OECD perform better on a number of dimensions such as economic growth, investment, capital stocks, and labour-force growth than their Catholic counterparts has been empirically demonstrated by Desdoigts (1999). In attempting to unearth the causes for the phenomenal success of East Asian economies, Kahn (1979) notes that a possible explanation may lie in the cultural commonalities rooted in the teachings of Confucius that these countries shared.

2.3.4 The institutions hypothesis

In recent times, institutions have been perhaps the most extensively researched cause of economic development. The view proposed by North (1989), that institutions are the primary cause of development, led to the birth of the new institutionalism/new institutional economics school of thought, which attributes the inability of many developing countries to escape the vicious cycle of poverty to poor institutions, which are themselves endogenous (Przeworski, 2004).

New institutionalism equates institutions to rules. This is clear from the following definition by North (1989):

> "Institutions are rules, enforcement characteristics of rules, and norms of behavior that structure repeated human interaction." (p. 1321)

Hence, the new institutionalist paradigm regards institutional reform as changes in the rules governing human interactions (H. J. Chang, 2011). Typically, the cost of measuring compliance to rules may be a major impediment to the effectiveness of institutions (North, 1989). This is further emphasised by H. J. Chang (2011), who notes that institutions are expensive to establish and maintain. This observation projects the presence of a reverse causality in the relationship between institutions and development, in that countries need to achieve a certain level of economic development before they can establish costly institutions like democracy, intellectual property rights, bankruptcy laws, security regulations, etc. Given that these institutions are strongly established in many developed countries, developing countries can establish superior institutions at a much lower cost, thereby enjoying a notable latecomer advantage (H. J. Chang, 2011).

The primacy of institutions in the development process is demonstrated by Rodrik et al. (2004), who estimate the contributions of institutions, trade and geography towards determining a country's income level. They show that when institutions are controlled for, the effect of geography on income is weak and trade has an insignificant effect on income. Acknowledging that the poor quality of institutions is a major obstacle to economic development, many international organisations and bi-lateral and multi-lateral donor schemes now impose conditions upon recipient countries, which requires them to improve the quality of their institutions so as to establish good governance (H. J. Chang, 2011). The Washington Consensus, for instance, was based on the view that all economies are inherently similar, and that, since people respond to incentives, a particular policy would yield the same results everywhere (Kenny & Williams, 2001). However, in practice, implementation of institutional reforms that have proven to be successful in one country/culture may not necessarily lead to desired results in another place. Emphasising this point further, Rodrik, Subramanian, and Trebbi (2004) note that "Western" institutional reforms need to be adapted to suit local conditions when they are implemented in developing countries. Przeworski (2004) considers the case of an independent judiciary in this regard; simply enforcing independence upon the judiciary would not lead to transparency if judges receive low salaries.

Much of the evidence on institutions as a primary cause of development is based on macro-level empirical analyses. However, the results of these studies need to be regarded with caution, as they are subject to aggregation errors, and they fail to take into account a number of micro-level institutions, such as land tenure systems, out-grower agreements and the management of individual projects funded by donor aid, which can influence human behaviours and exchange (Kimenyi, 2011).

Furthermore, the claim made by Acemoglu, Johnson, and Robinson (2005), that "differences in economic institutions are the fundamental cause of differences in economic development” has been challenged in some contexts. For instance, in the case of India, reforms such as industrial de-licensing, tariff reduction, financial liberalisation, and the promotion of small scale industries can only account for about a third of the 5% per annum increase in manufacturing productivity that occurred during the period 1993 to 2007 (Bollard, Klenow, & Sharma, 2013). Even more compelling is the evidence provided by Glaeser, La Porta, Lopez-De-Silanes, and Shleifer (2004) who demonstrate that human capital is a more fundamental cause of growth than institutions. Furthermore, as these authors, as well as numerous others have pointed out, there are several examples in recent economic history of countries that have achieved phenomenal economic growth under dictatorial regimes and subsequently adopted democracy and other institutions needed for good governance.

2.3.5 Technology and the elasticity of factor substitution as causes of economic growth

As technical change and the elasticity of factor substitution are considerations at the heart of the first essay, a brief survey of these factors as causes of economic growth is provided here. Neoclassical growth economists assume that economic growth primarily takes place due to advancements in technology and changes to the organisation of production activities (Hulten, 2001). Despite the key role played by technological progress, it is an exogenous process in the neoclassical growth framework (Mata & Louçã, 2009). Specifically, in the neoclassical model of economic growth proposed by Solow (1956) and Swan (1956), technology is perceived to be a public good that can be freely used by all countries (Fagerberg, 1994), and technological progress therefore occurs at the same rate all over the world (Kenny and Williams, 2001). If the extant pool of technology was indeed available to all countries, it would result in the absolute convergence of per capita incomes. However, the persistent gaps in income and growth rates between countries shed doubts on this assumption. The need to find explanations for these diversities led to the birth of endogenous growth theory that links these differences to processes relating to human capital accumulation and learning by doing. We provide a discussion of endogenous growth theories in Section 2.4.

With regard to the role of technology in economic growth, as noted previously, the Solow residual is often regarded to be a measure of technological change in an economy. However, in recent years, there has been increasing scepticism among many economists as to whether the Solow residual actually measures technological change. In relation to this concern, Lipsey and Carlaw (2000) distinguish between three alternative interpretations of TFP found in the extant literature: some economists accept TFP as a measure of technological progress; some believe that TFP is not a measure of technological progress, but is rather a measure of the free lunches emerging from scale effects and spill-overs, in that technological advancements enable people who are not directly involved in their creation to enjoy an improvement in income which is independent of the growth of output in the economy; finally, a third group doubts if it measures anything meaningful at all. Adding to this debate, Hulten (2001) points out that TFP does not only measure technological progress; it captures all of the factors commonly encompassed within a residual term in a regression equation such as measurement errors, model misspecification, and the effect of omitted variables. Hence, its value should be interpreted with much caution.

The institutional framework in a country is a primary determinant of technological progress. According to Boulhol (2004), institutions affect how efficiently technology is employed, the rate at which this efficiency improves, as well as the pace at which technology is diffused. Inventors and innovators will be unwilling to develop new technologies if they cannot reap profits from them. Hence, the presence of patents, copyrights, and other protection for inventors and innovators is a key determinant of technological progress. Overall, government policy plays a key role in determining the pace of technological development in a country, not only in relation to fostering innovation, but also with regard to directing firms to adopt the most appropriate and socially efficient technologies and negotiating better terms and conditions for adopting foreign technologies locally (Dahlman, Ross-Larson, & Westphal, 1987).

The view that the diffusion of new technologies is made possible through structural change was proffered by authors like Schumpeter (1934) and Chenery (1959). In particular, Schumpeter looked at a particular form of structural change that occurs when new industries gradually replace old ones, a process which he referred to as “creative destruction.” For instance, the industrial revolution created significant changes in the structure of economic activity and social organisation, and these changes in turn provided a foundation upon which new inventions, such as the steam engine and the spinning jenny, could be widely diffused. The *direction* of technical change is another influential strand in this body of literature, which was pioneered by Acemoglu’s (2002) paper, *Directed Technical Change*. Acemoglu (2002) makes a distinction between factor augmenting technical change and factor biased technical change. In a standard production function $f(K, L, A)$ where K denotes capital, L is labour and A is a technology index, technology augments, say capital, if the production function takes the form $f(AK, L)$. On the other hand, the technology parameter is biased towards capital if it satisfies the condition $\partial \frac{\frac{\partial F}{\partial K} \Big/ \frac{\partial F}{\partial L}}{\partial A} > 0$. The extent to which technology is biased towards one factor depends on the elasticity of substitution between factors. In the context of the CES production function, Acemoglu (2002) demonstrates that capital augmenting technical change is also biased towards capital when capital and labour are gross complements. However, when the two factors are gross complements, capital augmenting technology is actually biased towards labour. The explanation for the opposing directions of the two types of technical change under gross complementarity is that an increase in the productivity of one factor (due to technical change which augments this factor), causes the demand for the other factor to increase more than proportionately, which raises its marginal product.

Given its critical influence on the direction of technical change is the economy, the elasticity of factor substitution is generally considered to be a “deep” growth parameter (Acemoglu, 2002; Mallick, 2012a). Nonetheless, the question of how the elasticity of substitution between factors affects economic growth forms a burgeoning literature in its own right. Pitchford (1960) was one of the earliest studies dedicated to exploring this issue. In this paper, the author demonstrates that the economy’s growth rate and steady state level of output depends on the elasticity of factor substitution between capital and labour. Nevertheless, the increased interest in this issue in recent years emerged due to the findings of De La Grandville (1989) and Klump and De La Grandville (2000), who demonstrate that a country with a higher elasticity of factor substitution could always achieve a higher rate of growth in transition, as well as a higher level of output per capita at steady state. Adding to these findings, Klump and Preissler (2000) show that in a Solow-Swan model with a CES production function, a higher elasticity of factor substitution improves the potential for an economy to achieve sustained growth in the long run, and concurrently reduces the possibility of the emergence of poverty traps.

A number of empirical studies have revealed that the elasticity of factor substitution could indeed be a strong explanatory variable in cross country growth regressions. For instance, in a comparative study involving the US and South Korea, Yuhn (1991) shows that South Korea's higher elasticity of factor substitution was a key reason behind the impressive growth rates the country was able to register over several decades. In a more recent study, Mallick (2012b) uses data from 90 countries to show that the elasticity of factor substitution can account for about 20% of the variation in growth rates between Sub-Saharan Africa and East Asia.

2.3.6 Other causes of economic growth

There are some other causes of growth that have received considerable attention in the recent literature. Before concluding this section, we provide a brief overview of a few of these factors.

According to Lucas (1988), trade is also a major determinant of economic growth. He notes that the marked increase in exports was a key driver of the rapid economic development experienced, first by Japan, and later by South Korea, Taiwan, Hong Kong, and Singapore. The interest in trade and openness to international trade as determinants of economic growth occurred due to the diverse policies relating to international trade exercised by different regions during the latter part of the 20th century; on the one hand, Latin American countries followed a policy of import substitution, while on the other, East Asian countries actively pursued the outward-looking policy of export orientation. While Latin America lagged behind, the export orientation of East Asian economies enabled them to acquire technological knowhow from foreign firms quickly, which allowed them to achieve spectacular growth (Zhou, 2008). However, as Zhou (2008) notes, the export-led growth of East Asia occurred within a particular geopolitical structure that those countries shared, and openness to trade may not necessarily be a universal recipe for success.

Nevertheless, contrary to the view that free trade promotes growth, Yanikkaya (2003) provides empirical evidence that shows that developing countries may actually benefit from trade restrictions. Furthermore, Dollar and Kraay (2003) reveal that trade barriers have little impact on trade volumes in some countries. Hence, despite being an important dimension, there is no consensus in the literature that free trade is necessarily a primary reason for a country's development.

Migration is another important determinant of economic growth and development. The ways in which both temporary migration for purposes such as employment and education, as well as permanent migration affects growth in both the host and the source country have been extensively researched by growth/development economists over several decades. With regard to temporary migration, remittances by migrant workers account for over 1% of GDP in around 60 developing countries. Hence, remittances are considered to be a source of long run growth by a number of authors such as Ziesemer (2012), Barajas, Chami, Fullenkamp, Gapen, and Montiel (2009) and Chami et al. (2008). However, Rao and Hassan (2011) show that, through the contribution they make to the development of the financial sector, remittances may have an indirect rather than direct impact on long run growth in developing countries. Such a positive association between remittances and financial sector development is also shown by Aggarwal, Demirgüç-Kunt, and Pería (2011). Nevertheless, some empirical evidence suggests that the impact of remittances on economic growth is at best a modest one. For instance, the dynamic panel estimates obtained by Catrinescu, Leon-Ledesma, Piracha, and Quillin (2009) suggest that the effect of remittances upon long run economic growth is quite weak.

Some studies also suggest that remittances may affect growth negatively. For instance, Abdih, Chami, Dagher, and Montiel (2012) use a theoretical model and supporting empirical evidence to demonstrate that increased remittances could hinder growth by contributing towards a deterioration of institutional quality though two channels: firstly transfers received by households increase their consumption and thereby the tax base, which encourages government officials to misappropriate resources; secondly, these households are less likely to demand public services such as health care and education, as they can afford to pay for alternative services offered by the private sector, which reduces government expenditure on public goods and thereby encourages misappropriation. Further empirical evidence suggesting that remittances may hinder economic growth has been provided by, among others, Le (2009).

Permanent migration is also an issue that has enjoyed sustained attention in the growth literature over several decades. The emigration of educated professionals, more commonly known as brain drain, is a serious challenge that many developing countries face. Brain drain potentially has an institutional explanation because the poor state of democracy, the lack of independence of the judiciary and the poor quality of education and health care often motivate educated people to go in search of greener pastures. Nevertheless, the decision to emigrate is also primarily an economic one, because people make this decision with the expectation of earning higher wages abroad.

That permanent emigration may be detrimental for growth in the source country was the key outcome emulating from much of the earlier literature on the subject. These studies suggest various channels through which brain drain can adversely impact growth in developing countries. Some of the older studies dealing with this theme include Bhagwati and Hamada (1974), C. A. Rodriguez (1975a) and C. A. Rodriguez (1975b). Some recent theoretical works on this subject include a model developed by Di Maria and Stryszowski (2009), in which the expectation to migrate results in individuals in developing countries acquiring skills that are not appropriate for their country, which results in these countries being unable to converge to the technological frontier of their developed counterparts. In a study by Bénassy and Brezis (2013), the authors use a dynamic growth model with positive externalities to illustrate that, in a country where skilled workers form a small proportion of the population initially, their resulting wage will be lower, which encourages them to emigrate. Hence, such countries may become trapped in a history dependent poverty trap characterised by low human capital and high emigration. In accordance with these theoretical findings, the empirical evidence presented in Docquier, Lohest, and Marfouk (2007) shows that developing countries located close to OECD countries and those with colonial origins face a marked brain drain effect. Di Maria and Lazarova (2012) use a sample of developing countries to demonstrate that approximately 70% of the people in these countries suffer due to the adverse impact on growth resulting from the migration of skilled workers.

However, as noted by Lien and Wang (2005), there has recently been a notable shift in the literature towards highlighting the beneficial effects of permanent emigration. Along this vein, a number of theoretical studies such as Mountford (1997), Stark, Helmenstein, and Prskawetz (1997), Beine, Docquier, and Rapoport (2001), and Stark and Wang (2002) have argued that emigration, or the expectation to emigrate, may lead to a "brain gain" for the source economy. The mechanism suggested in these papers is roughly as follows: emigration prospects encourage people to invest in human capital, which results in a brain gain for the source economy. However, as not all of these individuals actually subsequently migrate, the extent of the brain drain is determined by the proportion of people who actually migrate. The source economy benefits from a higher average level of human capital and higher productivity, as long as the gain effect dominates the drain effect. Empirical evidence suggesting the presence of such a gain effect for developing countries is provided by Beine, Docquier, and Rapoport (2008) and Beine, Docquier, and Oden-Defoort (2011). In addition to the beneficial impacts on human capital development revealed in these studies, Skeldon (2008) notes that emigrants from China and Vietnam play a pivotal role in the development of their countries of origin through the large sums of money they remit and the active role they play in infrastructure development through the investments they undertake.

Many studies have also shown that developments in the financial sector are closely associated with economic growth. For instance, Roubini and Sala-i-Martin (1992) use theoretical and empirical evidence to demonstrate the adverse impacts on growth resulting from efforts made primarily by Latin American governments and monetary authorities in the 1970s to repress the financial sector, primarily by keeping interest rates artificially low and the strict state control of the banking sector in order to lower the cost of servicing government debt and exercise strict control over money supply. Using data for 47 countries covering the period 1976 to 1993, Levine and Zervos (1998) demonstrate that both stock market liquidity and the level of development of the banking sector display a robust positive association with capital accumulation and the growth of productivity and output. Odedokun (1996) use a sample of 71 developing countries to show that financial intermediation promotes economic growth in about 85% of the countries included in the study. De Gregorio and Guidotti (1995) also find a strong positive correlation between long run economic growth and financial development for a large cross country sample, and attribute this relationship to efficiencies in investments created through the development of the financial sector.

The mechanism through which the financial sector creates such growth promoting efficiencies is explained by Levine and King (1993) in the context of an endogenous growth model in which financial institutions screen and fund innovative entrepreneurs. These authors argue that as financial institutions are typically more efficient than individual investors at evaluating and monitoring projects, they contribute towards reducing the cost of productivity improvements, which in turn boosts economic growth.

2.4 HUMAN CAPITAL AS A SOURCE OF ECONOMIC GROWTH

Despite being regarded as a "proximate" cause of economic growth by Acemoglu (2008), human capital as a source of growth is a topic that has attracted persistent interest among academics and policymakers for several decades. As the second and third studies of the thesis explicitly deal with issues relating to human capital, we provide a brief overview of its definition, components, and relationship to economic growth in this section.

Research on human capital was pioneered by works such as Lewis (1954), Mincer (1958), Schultz (1961), Becker (1962), and Ben-Porath (1967). These authors emphasise that, unlike physical capital, human capital cannot be immediately obtained in the form of a final good. Ben-Porath (1967) notes that the creation of human capital is a task that an individual must undertake himself. Undertaking investments in human capital, typically in youth, causes an individual to forego potential earnings, but acquiring human capital enables her to enjoy higher earnings later in life. Such individual investments in human capital contribute towards improving the aggregate stock of human capital, which enables an economy to achieve a higher growth rate of output. However, such positive externalities are not taken into consideration by individuals, resulting in underinvestments in human capital (Stark, 2004).

Such positive externalities lie at the heart of endogenous growth theory. For instance, in Romer's (1986) model, long run growth is endogenously determined by the accumulation of knowledge. Later on, Lucas (1988) introduced a model in which long run economic growth was determined through human as well as physical capital. Generally, these authors used a production function of "AK" type to introduce non-diminishing returns to scale. By considering the term K to be a composite of both human and physical capital, these authors argued that the positive externalities emanating from investments in human capital, as well as physical capital, could generate sustained long run economic growth (Pack, 1994). Stemming from the attention paid to the centrality of human capital for economic growth by these pioneering authors, it has been the subject of large number of studies including Tamura (1991), Glomm and Ravikumar (1992), Benhabib and Spiegel (1994), Tallman and Wang (1994), and more recently Arcalean and Schiopu (2010), and Hanushek and Woessmann (2012).

Given the key role played by the individual in the creation of human capital, the production process associated with its formation is one that is non-standard (in comparison to the traditional production functions that depict the production of standard consumption goods), and characterised by considerable complexity. The plethora of factors governing the production of human capital is captured in the following definition of the human capital production function provided by Ben-Porath:

> "The technology which the individual faces when he makes decisions about investing in himself is a complicated system of technical and institutional relationships covering a wide spectrum of activities including formal education, acquisition of skills on the job, child care, nutrition, health, etc…" (1967, p. 359)

Much of the early work on human capital equated it to education and training. For instance, Mincer (1958) refers to it as training required to perform a particular job. Despite adopting such a narrow definition to facilitate simplicity of analysis, there has always been a consensus among researchers that the concept of human capital involves much more than education and on the job training. This is patently illustrated by Becker (1994) who notes that human capital refers to a person’s knowledge, abilities, state of health and values which, as opposed to physical or financial capital, cannot be separated from the owner.

However, Becker (2007) notes that when compared to the plethora of studies on human capital dealing with education and training, the studies focusing on health form a relatively small volume of literature. Nevertheless, as this thesis consists of explorations of both health and education, some further exploration regarding the importance of health as a form of human capital is believed to be warranted.

In the original framework for studying investments in human capital pioneered by economists like Gary S. Becker and Yoram Ben-Porath, improvements in knowledge and skills through investments in human capital entail a cost to the individual in terms of time and consumption expenditures. However, such investments improve a person’s market and non-market productivity, thereby enabling her to obtain a higher wage and produce more goods for household consumption. Although Becker (1964), for instance, claimed that his framework, which was developed with a particular focus on on-the-job-training, could be used to analyse investments in alternative types of human capital, Grossman (1972) argued that health capital is different from knowledge and skills in that it affects the amount of *time* a person can spend in income generating activities. He therefore maintained that investments in health cannot be studied within the traditional framework developed for studying human capital, and developed an alternative model for studying investments in health.

Delving further into the characteristics of health as a form of human capital, it comprises of the characteristics of both a consumer good and a producer good. Health is a producer good because it leads to improved labour productivity, which contributes towards higher economic growth, and in turn, improvements in technology (Stevens, 1977). However, good health is also an end in itself as it adds to a person's utility. For instance, in Grossman's (1972) model for the demand for health, this dual nature of health is captured by assuming that health affects the individual's preferences; poor health is a form of disutility. Concurrently, it acts as a producer/invesment commodity, because it is a determinant of the total number of days the individual might be unwell and unable to work, and as a result it affects his income. Health may affect growth directly or have an indirect impact on growth by influencing educational outcomes (Hazan & Zoabi, 2006).

Nevertheless, health and education are intrinsically related, as both are "joint investments made in the same individual" (Mushkin, 1962, p 130). The notion that they are joint investments might induce one to assume that the causality between health and education is bi-directional. However, the view that education leads to better health rather than the other way around seems to have emerged in the literature over time (see, for example, Mirowsky and Ross 1998, and Cipriani and Blackburn, 2002). Regardless of the causality of this relationship, they are both equally important forms of human capital. Perhaps the best embodiment of the equal importance of both health and education is the Human Development Index, which measures the progress of a country along three equally weighted dimensions: health, education, and income (Beraldo, Montolio, & Turati, 2009). The bulk of this thesis is grounded in the view that both of these types of human capital can contribute to economic growth and also determine how inequality in an economy evolves over time.

2.5 INEQUALITY: PATTERNS, CAUSES AND CONSEQUENCES

When exploring the distribution of income across countries, there are two dimensions of divergence that are often emphasised. The first is the idea of β-divergence, which captures the fact that the growth rates of poor countries have remained lower than those of their richer counterparts (Sala-i-Martin, 2002). The second is σ-divergence, which refers to the observation that inequalities in the distributions of incomes between countries have increased over time (Sala-i-Martin, 2006). However, inter-country income differences do not constitute the only dimension that is worthy of exploration in a discussion of inequality. Neither is income inequality the only type of inequality that deserves attention. Hence, in this section, we provide a brief discussion of inter and intra-country inequalities in income and wealth as well as the inequalities that prevail in other aspects such as human capital accumulation and land.

The dynamics of the *world income distribution* is a topic that has received substantial attention in the recent macroeconomics literature. According to Quah (1997), the world income distribution changed from a unimodal distribution in the 1960s to a bimodal one by the 1990s. Using data for the period 1960-1988, C. I. Jones (1997) also confirms this gradual polarisation of the world income distribution. As noted in Section 2.2, a twin-peaked global distribution of income supports the idea of convergence clubs, as countries cluster around one of the peaks, leading to the formation of two groups of countries—one rich and the other poor, which is associated with greater β-divergence. In contrast however, Acemoglu and Ventura (2002) suggest that the world income distribution has remained fairly stable since 1960. Although neoclassical growth literature attributes such stability to diminishing marginal products and technological spillovers, these authors suggest that international trade could also be a channel for stabilisation of incomes. Export prices decline in countries that accumulate capital faster, which results in a fall in the returns to capital, as a result of which capital accumulation in those countries reduces. However, such capital accumulation enables such countries to import more, which causes the returns to capital in other countries to increase, resulting in greater capital accumulation elsewhere, and the susequent decline in the returns to capital in those countries as well. These international trade induced diminishing returns to capital contribute towards stabilising the world distribution of income.

We now discuss how income distribution evolved in some individual countries. In the case of the USA, the distribution of income remained fairly stable from the 1940s to the 1970s. This relative stability of the income distribution resulted in researchers paying less attention to issues relating to inequality during this period (Gottschalk & Smeeding, 1997). However, from the 1970s to the 1990s, there was an exacerbation of income inequality in the US, which could be mainly attributed to differentials in earnings attached to education and experience.

Reiterating this observation, Picketty and Goldhammer (2014) point out that from the mid-1980s onwards, income inequality in the US rose rapidly, with the labour income earned by the top income decile rising from around 30-35% of national income in the 1970s to around 40-50% by the 2000s (p.298-299). Commenting further on the reasons for such an exacerbation of income inequality the the US, Goldin and Katz (2009) observe that the wage gap between graduates and workers with only a high school education diminished until the 1970s, only to record a steep increase after the mid-1980s. Around the same time that this change occurred, there was a slowdown in the number of people completing a university education. In a context where tertiary education is unaffordable to many, the continued demand for highly skilled workers coupled with stagnant supply caused a rise in the wages commanded by educated workers, and inequality increased as a result. Hence, the increase in income inequality in the US is primarily attributable to inequalities in the accumulation of human capital. Such experiences demonstrate the need to establish more egalitarian education systems.

A similar increase in income inequality was also seen in the UK, Canada, Japan, Germany, Sweden, and the Netherlands during this period (Gottschalk & Joyce, 1998). Furthermore, an increase in income inequality was oberved in Russia in the period following the introduction of free market reforms (Akhmedjonov, Lau, & İzgi, 2013). Such disparities in income, attributable to a large number of factors, such as differences in human capital, infrastructure, and attitudes of regional governing bodies, can also be observed between different states in India (see Ghosh, Ghoshray, & Malki, 2013, and references therein).

In contrast, however, between 1976 and 1993, the income Gini coefficient in South Korea fell by 27%, implying that miracle growth also concurrently led to a marked reduction in inequality (Fields & Yoo, 2000). The primary reason provided by these authors for this observed decline in income inequality is the reduction in the wage premium attached to higher educational levels and job tenure and the reductions in the gender wage gap. Such experiences point at the possibility that higher economic growth and lower income inequality can be achieved concurrently.

In a broad investigation of income distributions within and between countries, Milanovic and Yitzhaki (2002) decompose world income inequality into inter and intra-country components, and make overall comparisons between inter and intra-country inequalities for five continents: Asia, Africa, Western Europe, North America and Oceania, and Eastern Europe. In Asia, they find that inter-country inequality accounts for a greater proportion of overall inequality in the region than intra-country inequality, while in Latin America and Africa, the contribution made by inter-country inequality is quite small in comparison to the big intra-country inequalities that pervail. For the transition economies of Eastern Europe the two types of inequality seem to posit a similar contribution towards overall inequality. In the case of North America and Western Europe, both inter and intracountry inequalities are quite low.

Given that income inequality is inextricably linked to economic growth, the relationship between income inequality and economic growth is also an issue that has enjoyed sustained interest among development and macroeconomists for over half a century. One of the earliest explorations of the relationship between economic growth and income inequality was undertaken by Kuznets (1955), who observed an increase in the rate of economic growth and a concurrent reduction in income inequality in the United States, United Kingdom, and Germany since the 1920s. The Kuznets curve suggests that income inequality would increase during the early phases of development, remain stable for some time, and then start to decline as the economy reaches an advanced state of development.[5] The Kuznets curve paved the way to a large volume of literature dedicated to exploring the link between economic growth and income inequality. Acemoglu and Robinson (2002) reveal that there is evidence to support the presence of a Kuznets curve for a number countries including England, Sweden, Germany, France, Columbia, and Brazil. In contrast, there is evidence of a continuous decline in inequality in Norway, the Netherlands, South Korea, Japan, and Taiwan. In general, the literature is inconclusive about the relationship between

[5] Kuznets (1955) refers to this process as a “long swing” rather than an inverted U shape.

economic growth/development and income inequality; while some studies have identified an inverse U-shape, others have unearthed positive, negative, or inconclusive relationships (see Shin, 2012 for a detailed discussion).

Income inequality may be characterised by persistence. An early exploration of such persistent income inequality was undertaken by Kuznets (1955), who showed that in the United States, the top income decile undertook almost all of the household savings. He pointed out that higher savings would enable richer households to hold more income bearing assets, thereby potentially leading to a persistence of income and wealth inequality across time and generations. An interesting explanation for the persistence of such inequalities in income across generations is provided in a theoretical model developed by Durlauf (1996). In this framework, the neighbourhood parents select to live in exerts an important influence on the education, and hence, income of their offspring. Under the assumption that education funding is determined through a majority vote, the children of parents located in more affluent neighbourhoods that allocate more funds for education will become more educated and productive due to a positive "neighbourhood externality", while others will be less educated, resulting in the persistence of income inequality across generations. However, the persistence of income inequality may be the result of government policy.

The initial distribution of income can be a major determinant of how poverty and inequality in an economy evolves over time (Ravallion, 1997). However, Deininger and Squire (1998) empirically demonstrate that the initial distribution of *land*—which can be a proxy for the initial distribution of assets—is more strongly related to subsequent growth than the initial distribution of income. In particular, they show that an initial unequal distribution of assets reduces the economy's ability to achieve subsequent growth. They provide two potential explanations for this observation.

Firstly, agents possessing less assets are unable to obtain credit to make economically viable investments in activities such as education, which is evident from the observation that high initial land inequality leads to lower subsequent educational attainments among the population of a country. The other is a political economy explanation, whereby an individual's holdings of assets affects her ability to participate in the voting mechanism. This is also evident from the observation made by the authors that in democratic regimes, initial asset inequality is not a significant determinant of subsequent growth.

Much of the recent growth literature has focused considerable attention towards providing explanations for the similar persistence of inequalities in human capital across generations. Given the broad definition of human capital adopted throughout this thesis, persistence of human capital inequality could occur when the offspring of educated parents acquire higher levels of education, while offspring of uneducated parents remain uneducated themselves, or when good/poor health persists across generations. The extant literature suggests numerous mechanisms through which such persistence of inequalities in human capital manifest themselves. For instance, Castelló-Climent and Doménech (2008) demonstrate that low parental human capital reduces the life expectancy of their offspring. Their lower lifespan discourages the children of poor agents from investing in education, as a result of which these dynasties remain trapped in a state of poverty. Another explanation for persistent human capital inequality is provided by Galor and Zeira (1993), who suggest that in the presence of indivisibilities in human capital investment combined with credit market imperfections, wealth inequality affects human capital investment and thereby economic growth in the long run.

A number of studies have also suggested that the initial distribution of wealth in a country may be the primary cause of such persistent inequalities in human capital. Studies such as Glomm and Ravikumar (1992), Ray and Streufert (1993), Deaton (2003), and more recently, Chakraborty and Das (2005), Das (2007), Sarkar (2008), and McDonald and Zhang (2012) demostrate that inequalities in the distribution of income and wealth could lead to unequal investments in human capital, which in turn could lead to a persistence of income and wealth inequality in across generations.

In particular, inequalities in wealth or human capital can affect an economy's growth trajectory through the political economy process. The political economy process is the means by which the interests of different individuals in society can be incorporated into policy decisions (Persson & Tabellini, 1994). The preferences of individuals regarding policies are typically often dependent upon their income, wealth, or human capital. Hence, the distribution of income and wealth in the economy could be a major determinant of political economy decisions upon which the economy's long run outcomes hinge. For instance, Alesina & Rodrik (1994) develop an endogenous growth model in which taxation on capital is determined through majority voting to demonstrate that under a more equitable distribution of capital the median voter would be better endowed with capital, and a lower rate of capital taxation prevails in the economy, which accelerates economic growth.

Whether inequality supports or hinders growth is a question that has attracted considerable attention in the growth literature. There is mixed evidence in both the theoretical and empirical growth literature regarding this direction of causality. For instance, a notable theoretical contribution in this regard is the study by Galor and Moav (2004), where they use a unified growth model to show that in the early stages of development, when the accumulation of physical capital drives economic growth, wealth inequality leads to a rise in savings, which translates into the accumulation of physical capital, thereby leading to higher economic growth. However, given a complementarity between capital and skills, this expansion of the capital stock encourages investments in human capital. In the later stages of the development process, when human capital becomes the key driver of growth, however, a more equal distribution of income and wealth is needed to encourage credit constrained individuals to undertake optimal investments in human capital. Hence, their theory argues that in the early stages of development, inequality is conducive to growth, while in the later stages, when human capital replaces physical capital as the primary engine of growth, inequality may impede growth. Along a similar vein, Galor and Tsiddon (1997) observe that a child's human capital is positively related to parental human capital, which they refer to as a home environment externality, while an economy's technological process is accelerated by the average level of human capital, a process they refer to as the global technological externality. If the home environment externality dominates the global technology externality, polarisation of the distribution of human capital occurs, as a result of which the segments of society that initially possess higher levels of human capital undertake high investments in human capital, while those dynasties with lower stocks of human capital invest less on acquiring human capital. However, at the aggregate level, the economy's average stock of human capital rises and this induces higher growth. This increase in the average stock of human capital ultimately causes the global technology externality to become dominant. By improving the returns to human capital, this reversal of impacts encourages the poorer members of society to invest more in human capital, ultimately leading to long run equality and prosperity.

However, some studies project the more pessimistic view that inequality may be a persistent, irreversible process. For instance, one argument suggests that in the presence of imperfect credit markets, the poor will not be able to borrow to invest in growth promoting activities. Given that the marginal product of capital is decreasing, this leads to a greater loss in output of the poor agents in an economy, thereby exacerbating inequality further and retarding growth (Ravallion, 2001). Empirical evidence in support of this inverse relationship between inequality and growth has been provided by, among others, Clarke (1995), Alesina and Perotti (1996), Birdsall and Londono (1997), and Lopez, Thomas, and Wang (1998).

2.6 CONCLUSION

This chapter provided a broad overview of some of the themes of direct and tangential relevance to the three studies that constitute this thesis. We looked at some cross country patterns of economic growth, explored a variety of causes of economic growth that have been suggested in the extant literature, looked at inequality primarily along the dimensions of income, wealth and human capital, and briefly discussed the intricate links between growth, human capital, and inequality. While this chapter is not meant to be a comprehensive survey of these macroeconomic issues, it provides an overall motivation of the ideas presented in the rest of the thesis.

Chapter 3: Technical Change, Variable Elasticity of Substitution and economic growth

3.1 INTRODUCTION

The elasticity of substitution between labour and capital is an important determinant of an economy's long run performance. Ever since de La Grandville (1989) used a CES production function to demonstrate that a country with a higher elasticity of substitution could achieve a higher rate of economic growth and a higher value of income per capita at steady state, there has been a considerable increase in the volume of research, both theoretical and empirical, on this theme (see, for instance, Karagiannis et al., 2005; Klump & De La Grandville, 2000; Klump & Preissler, 2000; Mallick, 2010, 2012b; Miyagiwa & Papageorgiou, 2007; Palivos & Karagiannis, 2010; Yip & Xue, 2013; Yuhn, 1991).

In what was perhaps the earliest empirical study on the elasticity of factor substitution, Arrow et al. (1961) observed that factor substitutability can vary between industries and across countries. More recently, this view has been expounded by a number of studies that have concluded that the elasticity of factor substitution, typically assumed constant in standard growth models, could instead be an endogenous variable related to the level of economic development of a country (see, for example, Miyagiwa & Papageorgiou, 2007). For instance, Duffy and Papageorgiou (2000) use cross-country data to show that generally, richer countries have an elasticity of factor substitution above 1, while poorer countries have an elasticity of factor substitution below 1. Most extant studies that have explored the way in which the elasticity of factor substitution could impact on the long run behaviour of the economy have used the CES production function, first discussed in Solow (1956). In the CES production function, the value of the elasticity of substitution remains constant across different input combinations and over time. Hence, it cannot take into account the variations in the elasticity of substitution estimates revealed in the studies referenced above.

The variety of estimates for the elasticity of factor substitution available in the CES literature is suggestive of a possible mis-specification of the underlying production function, which creates the need to consider alternative, more flexible production functions that can capture the differences in factor substitutability between developed and developing countries and between different industries that are revealed in the extant literature. In this regard, the variable elasticity of substitution (VES) production function, first discussed by Sato and Hoffman (1968) and Revankar (1971), is an appealing alternative to the CES production function. Much of the properties and dynamics associated with the VES production function are determined by a technology parameter labelled b. The value of this parameter is a crucial determinant of a number of key features of the economy such as capital intensity, the direction of technical change, and the share of capital in output. Through the technical change it generates, this parameter relates the elasticity of substitution between labour and capital linearly to the capital stock per worker. Positive values of b result in an elasticity of substitution above 1, while negative values lead to an elasticity of substitution below 1. Thus, the elasticity of substitution is endogenous, in that it is driven by the parameter b and the economy's capital-labour ratio. Differences in the value of this parameter may therefore provide a potential explanation for the empirically observed differences in the elasticity of factor substitution between countries and industries.

The empirical relevance of the VES form has been demonstrated in a number of studies that have suggested that the VES production function can be a better specification than the CES production function. For instance, in the case of urban housing construction, where there are considerable variations in the ratio of land to non-land inputs, authors such as Sirmans, Kau, and C. F. Lee (1979) and Färe and Yoon (1981) show that assuming a VES production function allows the elasticity of substitution between land and non-land inputs to depend on relative land intensity, and results in a better specification when compared to other forms of production function. Using data for Japan covering the period 1878-1938, Bairam (1989) demonstrates that the capital-labour ratio, as well as the elasticity of substitution between capital and labour increased during this time, suggesting that the VES form is the correct functional specification of the production function for the Japanese economy during this period. Using data on Indian manufacturing industries, Kazi (1980) demonstrates that the elasticity of factor substitution varies between industries, and that assuming a CES production function leads to an upward bias of the estimate of the elasticity of substitution, and thereby concludes that the VES production function is the correct functional specification to account for these variations in factor substitutability.

Despite the appealing properties and empirical relevance of the VES production function, it has received limited attention in the theoretical growth literature. To the best of our knowledge, within the theoretical literature exploring the long run macroeconomic outcomes associated with the elasticity of substitution, there have been only two studies that employed the VES production function. The first is by Karagiannis et al. (2005), who incorporate a VES production function into an otherwise standard Solow-Swan model and demonstrate that an economy can display unbounded growth if the elasticity of factor substitution exceeds a certain minimum value. The other is a paper by Brianzoni, Mammana and Michetti (2012), who consider the VES production function in the context of a discrete time one sector Solow–Swan growth model with differential savings and demonstrate that small values of the elasticity of substitution, resulting from negative values of the parameter b, could result in the emergence of complex dynamics.

In view of the interesting features and empirical relevance of the VES production function, the first aim of this paper is to explore the properties of the VES production function in greater depth. This exercise enables us to discuss the properties of the VES production function in the context of a number of stylised features of economic growth that have been observed in the recent literature. Secondly, the VES production function is applied to an otherwise standard Diamond (1965) overlapping generations (OLG) model. This analysis is warranted because extant theoretical applications of the VES production function mentioned in the preceding paragraph apply it in the context of the Solow-Swan model. The overlapping generations model, by taking into account the life cycle aspects of a consumer's utility maximising problem, possesses the ability to provide potentially deeper insights into long run outcomes and transitional dynamics than the Solow-Swan model. In contrast to the literature on the CES production function, which concentrates on the various macroeconomic effects associated with the elasticity of substitution, this paper focusses on the effects upon the long run behaviour of the economy created by the technology parameter b and investigates a hitherto unexplored aspect of the VES production function, which is the manner in which the *technical change* generated by this parameter impacts on the *mechanisms of growth*.

This exploration of the properties of the VES production function reveals a number of interesting results. It demonstrates that the technology parameter b influences the marginal products of capital and labour capital differently, with the marginal product of capital rising in the value of b, while the marginal product of labour is falling in it. By influencing relative marginal products, the parameter b determines the direction of technical change in the economy. A higher value of b leads to capital biased technical change by raising the relative marginal product of capital.[6] We are also able to show that a higher value of the parameter b is associated with steeper long run expansion path, implying that it leads to a higher capital intensity in production.

The VES production function is also a flexible functional specification that can potentially provide an explanation for the movements in factor shares experienced by many countries by tracing them back to changes in the technology parameter b. Figure 3.1 below shows that the labour share in output has displayed considerable movements in different countries. Countries such as the United Kingdom, France, and Australia have experienced an upward trend in the labour share in recent years, while nations that were significantly affected by the global financial crisis, such as Greece and Ireland, have seen noticeable declines in labour share in recent times.

[6] As per the definitions presented in Acemoglu (2002), capital biased technical change is identified by a rise in the marginal product of capital to relative to the marginal product of labour. Ceteris paribus, this would alter the optimal input mix towards the utilisation of more capital. In equilibrium, however, outcomes depend not only on this ratio (which is the marginal rate of technical substitution), but also the ratio of prices of these two inputs.

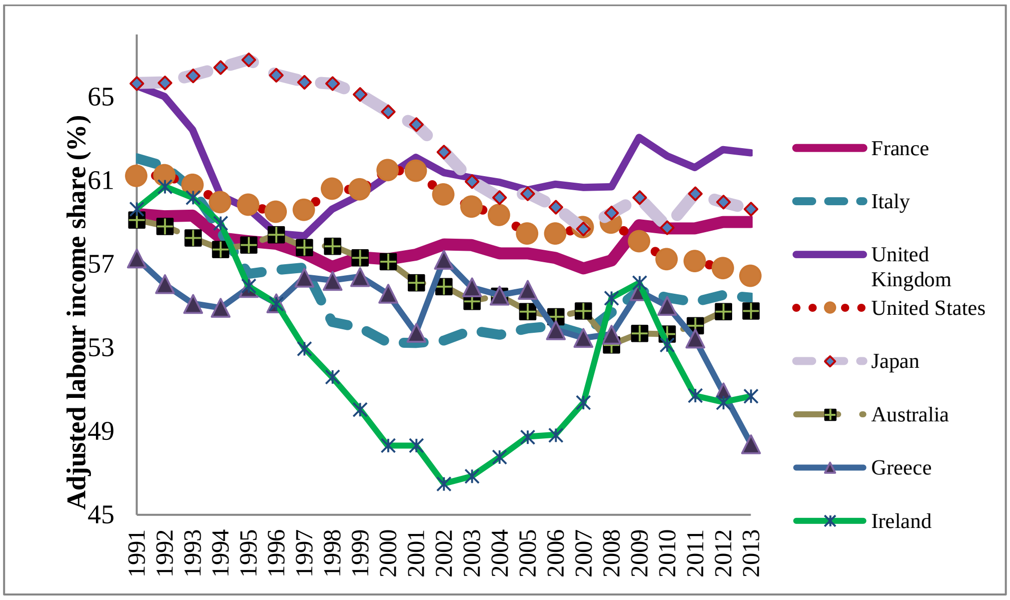


Figure 3.1: Movements in the labour shares in selected countries since 1991

Source: ILO (2014)

Noticeable movements in aggregate factor shares within and across countries such as those seen in Figure 3.1 have been observed by a number of authors (see, for instance, Acemoglu, 2003; Boldrin, 1995; Durlauf, 1995; Gollin, 2002; Solow, 1997; Young, 2010). However, despite evidence of such diversity in the movements of labour shares, recent research such as Karabarbounis and Neiman (2014), F. Rodriguez and Jayadev (2010), ILO (2013), and OECD (2012b), that cover a large number of countries, suggests that there has been a global decline in labour share.[7] [8] Adding to such observations, Piketty and Goldhammer (2014) warn of the extent of the potential increases in the share of capital of income by noting that, "It is quite possible that

[7] For a current and comprehensive discussion about the declining trend in the global labour share, we direct the reader's attention to pages 67-74 of Karabarbounis and Neiman (2014).

[8] The findings of these recent studies are in contrast to the observations in Acemoglu (2003), who uses US and French data for the periods 1929-1997 and 1913-1997 and argues that, despite considerable movements, factor shares have remained fairly constant in these countries. The assumptions of a Cobb-Douglas production function and labour augmenting technical change are both consistent with Acemoglu's observations. However, the more recent evidence presented in the studies cited above motivates a further exploration of these issues.

capital's share will increase in coming decades to the level it reached at the beginning of the nineteenth century." (p. 223). In the VES production function, a higher value of the parameter b is associated with a declining share of labour in output. Hence, the VES form provides a possible explanation for the decline in labour share experienced by many countries by attributing it a rise in the value of the parameter b.[9]

Upon exploring these features of the VES production function, we proceed to apply it to a Diamond OLG model. We are able to show analytically that for positive values of the technology parameter b, which lead to elasticity of substitution which is greater than 1, the economy always reaches a unique and stable steady state. When $b<0$, which results in an elasticity of factor substitution that is between 0 and 1, the economy could either be characterised by no non-trivial steady state, or two steady states, of which one is stable and the other is not. However, when $b<0$, the range of values of the capital stock per capita in which the VES production function displays standard properties of a production function is bounded from above. Hence, under the assumption of a VES production function, the Diamond model could display three possible growth trajectories: monotonic convergence towards a stable steady state, the emergence of a poverty trap, or a movement towards the upper bound of the per capita capital stock.[10] Therefore, in our model, conditional convergence towards a unique

[9] The literature referenced in the paragraph attributes the decline in labour share to technological change, the falling price of investment goods, increased globalisation, developments in the financial sector, and the deterioration of labour market institutions and the welfare state. However, the assumption of a VES production function explains the falling share solely in terms of the capital biased technical change captured by a higher value of the parameter b.

[10] Miyagiwa and Papageorgiou (2003), who apply a normalised CES production function to an otherwise standard Diamond model, show a similar result. However, in their model, the dynamics are directly linked to the (constant) elasticity of substitution between capital and labour, and hence, entail an interpretation that differs from our results. In our model, the dynamics are driven by the biased technological change created by the parameter b.

steady state like that observed in the context of a standard Solow-Swan model is not inevitable.

We also prove that the steady state capital stock is falling in the technology parameter b, but that steady state output is positively related to it. Hence, a higher value of the parameter b is associated with greater productivity at steady state. This outcome demonstrates that although a higher value of b is associated with greater capital intensity, the equilibrium quantity of capital per worker used is determined by marginal products, as well as market prices. Hence, capital-biased technical change does not necessarily imply that more capital will be used in equilibrium. In fact, in our model, capital biased technical change is manifested at steady state through the ability of the economy to produce a higher output with less capital per worker.

In the presence of competitive factor markets, the reward for each factor is equal to its marginal product. Given that a higher value of the technology parameter b is associated with a higher marginal product of capital and a lower marginal product of labour, it leads to an exacerbation of inequality. This worsening of inequality is an artefact created by the heterogeneity intrinsic to the Diamond model, viz factor shares associated with capital and labour are owned by different agents—the old and the young respectively. The benefits of capital biased technical change generated through a higher value of b then accrue to the old, who are the owners of capital.[11]

Rather than merely exploring the link between the elasticity of factor substitution and economic growth that has been the central objective of many extant studies, the aim of this paper is to explore how the parameter b impacts on the optimal capital

[11] According to Irmen and Klumpp (2009), a higher elasticity of substitution is also likely to lead to higher intergenerational inequality in the case of the CES production function, although the result can be more explicitly demonstrated in the context of a VES production function.

intensity in the economy, the rewards and relative marginal products of factors, the elasticity of substitution between labour and capital, and thereby the economy's transition dynamics and productivity in the long run. Such a shift in focus is necessitated firstly by the fact that b has a direct impact on the direction of technical change in the economy, and secondly because the elasticity of factor substitution is an endogenous variable driven by this parameter. As such, our study is related to the idea of directed technical change put forward by Acemoglu (2002). The use of the VES further provides the advantage of greater tractability relative to CES, in addition to the flexibility implied by the endogeneity of the elasticity of substitution. Furthermore, the explicit exploration of how changes in factor substitutability stemming from the biased technical change created by the parameter b can impact on intergenerational inequality also contributes to an aspect that has received sparse attention in the extant literature.

The rest of the essay is organised as follows: Section 3.2 provides a discussion of the VES production function and its properties, the model and analysis is provided in Section 3.3 and the chapter concludes with Section 3.4. A number of proofs and derivations are given in the Appendices.

3.2 THE VES PRODUCTION FUNCTION AND ITS PROPERTIES

The VES production function was introduced by Sato and Hoffman (1968) and Revankar (1971). Many of the properties of the VES production function have been established by Karagiannis et al. (2005). We first provide a detailed discussion of these properties before moving on to discuss the dynamics created by the VES production function in the context of a Diamond overlapping generations model.

In this paper the form of the VES production function proposed by Karagiannis et al. (2005), which is expressed as:

$$Y = AK^{a\nu}\left(L + abK\right)^{(1-a)\nu} \tag{3.2.1}$$

In (3.2.1) above, $\nu > 0$ is the returns to scale parameter and $A > 0$ captures Hicks-neutral technological change. In the remainder of the paper, we assume that the production function displays constant returns to scale so that $\nu = 1$ and we also assume without loss of generality that $A = 1$. Being the exponent of K, the parameter a can be considered as the "pure" share of capital in the production function. The parameter b, as we will see shortly, is an important technology parameter that impacts on the direction of technical change in the economy.

The VES form embeds three special cases. In equation (3.2.1), when $a = 1$, the VES production function takes the Harrod-Domar fixed coefficients form, more commonly known as the AK form, as output is linearly related to the stock of capital. Similarly, when $a = 0$, output is linearly related to the quantity of labour. Finally, when $b = 0$, the VES production function takes the Cobb Douglas form, and for this reason, it is regarded as a generalised form of the Cobb Douglas production function.[12]

The intensive form of the production function, where $y = Y/L$ and $k = K/L$, can be expressed as follows:

$$y = f(k) = k^{a}\left(1 + abk\right)^{1-a} \tag{3.2.2}$$

The parameters must conform to the range $0 \le a \le 1$, and if $b < 0$, the per capita capital stock is bounded from above such that $k < 1/|b|$. These restrictions are necessary to ensure that the VES form is characterised by the standard properties of a

[12] The fact that this particular form of the VES production function is a Cobb Douglas generalisation adds to its analytical tractability, and is one of its most appealing properties (Bairam, 1989).

production function. First note that $f'(k) > 0$ and $f''(k) < 0$ are standard properties of the intensive form production function. These derivatives are as follows:

$$f'(k) = \frac{a(1+bk)}{k^{1-a}(1+abk)^a} \tag{3.2.3}$$

and,

$$f''(k) = a(a-1)k^{a-2}(1+abk)^{-a-1} \tag{3.2.4}$$

From (3.2.3) and (3.2.4) above, we can see that the standard properties of a production function, which are $f'(k) > 0$ and $f''(k) < 0$, are satisfied only when $0 \leq a \leq 1$.[13] Furthermore, for the condition $f'(k) > 0$ to be satisfied when $b > 0$, we need $k > -1/b$, which is always satisfied since the capital stock in any given period must be real and positive. When $b < 0$, this condition is satisfied if $k < 1/|b|$.[14]

Note that (3.2.3) above is the marginal product of capital (MP_K). By differentiating (3.2.1) with respect to *L*, we get the marginal product of labour (MP_L) which is:

$$MP_L = \frac{\partial Y}{\partial L} = (1-a)\left(\frac{k}{1+abk}\right)^a \tag{3.2.5}$$

---

[13] Although the inequality signs should technically be strict in this instance, as explained below, we include equality in order to incorporate the special cases of perfect substitutes and the Harrod-Domar fixed coefficient.

[14] Note that this range is different to that given in Karagiannis et al. (2005), who do not take into account the effect on the inequality created by the negative sign in front of *b*. Revankar (1971), on the other hand gives the range correctly, although it is expressed differently from ours.

The value of b affects the marginal products of labour and capital differently. From equations (3.2.3) and (3.2.5) we have:

Lemma 1: A higher value of b is associated with a lower marginal product of labour but a higher marginal product of capital.

While it is clear from (3.2.5) that that MP_L is monotonically decreasing in b, we show in Appendix A that MP_K is always increasing in b. Since MP_K is rising in b and MP_L is falling in b, a higher value of b is associated with a steeper isoquant. As will be demonstrated shortly, this has important implications for the direction of technical change, as a higher value of the parameter b is associated with capital biased technical change. In a competitive market, the price of each factor will be equal to its marginal product. Thus, $MP_L = w$ and $MP_K = r$, where w and r are the wage and interest rates respectively. Therefore, a higher elasticity of substitution, caused by a higher value of b is associated with a lower reward for labour and a higher reward for capital. As mentioned in the Introduction, this observation has important implications for intergenerational inequality in the context of a Diamond model, where the stock of capital is owned by the old generation while the young generations supply labour in return for a wage. This is a consideration we will return to in Section 3.3.

The elasticity of substitution between capital and labour, which we denote with σ, is given by the formula $\sigma = \dfrac{d\ln\left(K/L\right)}{d\ln\left(MP_K/MP_L\right)}$. In the VES production function the elasticity of substitution between capital and labour is:

$$\sigma = 1 + bk \tag{3.2.6}$$

From equation (3.2.6) above, we can see that when $b > 0$, the condition $k > -1/b$ must be satisfied to ensure that $\sigma \geq 0$. On the other hand, when $b < 0$, the required condition is $k < 1/|b|$. These conditions are analogous to those identified earlier as necessary for a positive first derivative.

Note that the value of σ rises with the capital-to-labour ratio if $b > 0$ and falls if $b < 0$.[15] Therefore, in the VES production function, the value of σ changes along a given isoquant. As the capital-labour ratio is constant along a particular ray from the origin drawn through a map of isoquants, this feature implies that, for a particular value of b, the elasticity of substitution will be identical only at points along a ray from the origin (Revankar, 1971). This feature is in contrast to the CES production function in which σ remains constant at all points on an isoquant map. To illustrate these features, in Figure 3.2 below, we plot the isoquants associated with different values of b when the value of the parameter a is 0.5, given an output of 10 units. From the intensive form of the VES production function given in equation (3.2.2) it is evident that output is monotonically increasing in the value of b. In terms of the isoquant, this implies that a higher value of b enables the firm to produce a particular level of output with less inputs, resulting in the isoquant shifting inwards as the value of b rises, as shown in Figure 3.2.[16] This inward shift of the isoquant resulting from a higher elasticity of

---

[15] This strict monotonicity in the behaviour of σ with respect to capital intensity is considered to be a shortcoming of the VES production function (Bairam, 1989).

[16] In the CES literature, the technique of normalisation, which was introduced by de La Grandville (1989), is usually employed to ensure that isoquants relating to different values of the elasticity of substitution share a common point of tangency, which allows one to analyse the pure effects of the elasticity of substitution upon growth. No normalisation of the VES production function is carried out in this paper, because, rather than a narrow focus on the dynamics created by the elasticity of substitution, our objective is to explore the dynamics of growth generated by the parameter b, which influences a number of features of the economy, of which the elasticity of substitution is only one.

substitution is a characteristic the VES and the CES production functions share in common.[17]

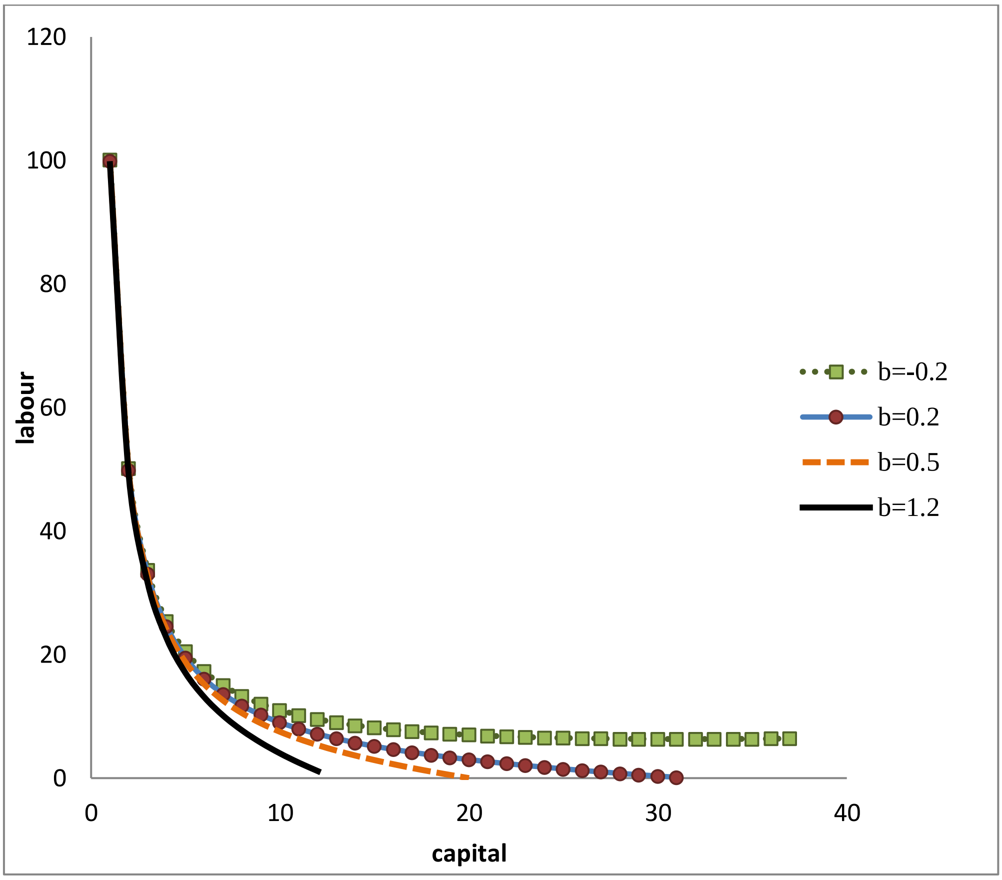


Figure 3.2: VES Isoquants associated with different values of b given $Y = 10$ and $a = 0.5$

Another interesting feature of the VES production function is that capital is the essential input in the production process: i.e. $Y = 0$ when $K = 0$. On the other hand, when $b > 0$, even when $L = 0$, a positive level of output can be produced as long as a positive quantity of capital is employed in production. As is evident from Figure 3.2, this feature results in the isoquant having an intercept on the capital axis when $b > 0$.

[17] Extant theoretical studies utilising the VES, such as Karagiannis et al. (2005) and Brianzoni et al. (2012) do not use any normalisation technique in their studies either. Note that the particular variant of the VES production function we consider is a Cobb-Douglas generalisation. CES generalisations of the VES production function do exist (see, for instance, Kadiyala, 1972; Lu & Fletcher, 1968), although they have remained in relative obscurity in both the theoretical and empirical spheres, possibly due to their complicated nature. Normalising such production functions may be possible, and is a potentially fruitful avenue for future research in the area.

While assuming one input is essential may appear restrictive, interpreting our capital term as a composite input comprising of both human and physical capital like in Romer (1986) is more plausible from a practical point of view.

However, as demonstrated by Karagiannis et al. (2005), this essential input feature causes one Inada condition to be violated.[18] As shown by these authors, the limiting properties of the VES production function are as follows:

When $b > 0$,

$$\lim_{k\to 0} f(k) = 0, \lim_{k\to\infty} f(k) = \infty, \lim_{k\to 0} f'(k) = \infty \text{ and } \lim_{k\to\infty} f'(k) = A(ab)^{1-a} \quad (3.2.7)$$

When $b < 0$,

$$\lim_{k\to {}^{1}/_{|b|}} f(k) = \frac{A(1-a)^{1-a}}{|b|^a} \quad \text{and} \quad \lim_{k\to {}^{1}/_{|b|}} f'(k) = \frac{A(1-a)^{1-a}}{|b|^a} \quad (3.2.8)$$

Hence, the per capita capital stock in any given period is strictly bounded from below when $b < 0$. The essential input feature of the VES production function causes one of the Inada conditions to be violated, as it means that even when labour approaches zero, the economy can produce a positive level of output.

We have already seen that the parameter b is instrumental in determining the elasticity of substitution between capital and labour in the VES production function. As mentioned in the Introduction, in contrast to the CES literature, where the discussions are generally centred around the elasticity of substitution σ, in this study, we shift our attention to b. In order to understand the mechanism underlying the influence the parameter b exerts on the value of the elasticity of substitution, the

[18] This result is further confirmed by Irmen and Maußner (2014) who establish that the Inada conditions hold only when both inputs in the production function are essential.

manner in which b affects the direction of technical change in the economy must be explored. For this purpose, the shares of labour and capital S_L and S_K are given below:

$$S_L = \frac{wL}{Y} = \frac{w}{y} = \frac{(1-a)}{1+abk} \tag{3.2.9}$$

$$S_K = \frac{rK}{Y} = \frac{rk}{y} = \frac{a+abk}{1+abk} \tag{3.2.10}$$

It is evident from equation (3.2.9) that a higher value of the technology parameter b causes the share of labour to fall. On the other hand, from equation (3.2.10), the derivative of the capital share with respect to the parameter b is given by $\frac{ak(1-a)}{(1+abk)^2}$, which is always positive. As mentioned in the Introduction, recent growth literature suggests a global decline in the share of labour in total output. Therefore, in terms of the VES production function, such a decline in the share of labour, and a concurrent increase in the share of capital, can be attributed to an exogenous rise in the value of the technology parameter b.

After having looked at how factor shares are affected by b, we next note that capital intensity is not fixed. Rather, it is optimally chosen by a firm or economy conditional on input prices and the production plan. To look at how the value of b influences capital intensity, we obtain the economy's long run expansion path by minimising the cost function associated with producing a certain level of output Y. Thus, for given factor prices w and r, we minimise total cost:

$$C = wL + rK \tag{3.2.11}$$

Subject to the constraint:

$$Y = K^{a}\left(L + abK\right)^{1-a} \tag{3.2.12}$$

Solving this problem yields the following long run expansion path:

$$\frac{K}{L} = \frac{\frac{w}{r}}{\frac{1-a}{a} - \frac{w}{r}b} \tag{3.2.13}$$

We can see from (3.2.13) that a higher value of the parameter b is associated with a higher capital to labour ratio and thereby a steeper long run expansion path. In other words, a higher elasticity of substitution yields greater capital intensity in production. This result can be expected because it was shown earlier that a higher value of b causes MP_K to rise and MP_L to fall. Hence, a higher value of b naturally encourages firms to substitute labour with capital, which is the relatively more productive input. Furthermore, as noted previously, capital is the essential input in the production process. Therefore, we observe that as the value of b rises, firms prefer to raise the quantity of capital, which is also the essential input in the production process.

We have seen that the value of the parameter b is a key determinant of the optimal capital-to-labour ratio/capital intensity in this economy. Hence, the parameter b influences the value of the elasticity of substitution in two ways: directly, by appearing on the right hand side of equation (3.2.6) above, as well as indirectly, by acting as a determinant of the capital per worker/capital intensity k. Thus, a higher value of b, while leading to a higher elasticity of substitution in its own right, also exerts a secondary incremental effect on the elasticity of substitution by contributing towards greater capital intensity.

It is also worth commenting on the fact that higher value of the parameter a is also associated with a higher capital to labour ratio. The greater capital intensification associated with this parameter lends support for our earlier interpretation of it as the "pure" share of capital in the production function.

Acemoglu (2002) suggests that under the assumption of a CES production function, even in the presence of an abundant supply of capital, as long as the elasticity of factor substitution is above a threshold value which falls between 1 and 2, the rewards to capital would rise, leading to capital-biased technical change. To explore what connotations the VES form holds for the direction of technical change, we define the factor bias as per Acemoglu (2002), who defines it as the ratio of marginal products. Hence, for the VES form, the capital bias, denoted by γ, is given by:

$$\gamma = \frac{MP_K}{MP_L} = \frac{a}{(1-a)}\left(\frac{1}{k} + b\right) \tag{3.2.14}$$

From equation (3.2.14), it is clear that γ is falling in the capital intensity k, as would be expected due to the substitution effect, which yields a negatively sloped relative demand curve for capital (Acemoglu, 2002), but it is rising in the parameter b. Hence, equation (3.2.14) demonstrates that a higher value of the parameter b leads to capital biased technical change, which can be interpreted as a situation where an exogenous increase in the value of b enables an economy to produce more output with the same amount of capital. This has some interesting and profound implications upon the long run behaviour of the economy, which are explored in the next section, where we present the model and discuss its dynamics.

Many extant studies assume that the elasticity of substitution is an exogenous determinant of the technical change in the economy (Mallick, 2012b), and only a few studies, such as Miyagiwa and Papageorgiou (2007) and R. W. Jones (2008), have explored the issue of endogeneity of the elasticity of factor substitution. The VES production function connects with the latter view, as the elasticity of substitution is a result of the directed technical change created by the value of the parameter b. Further to this point, it often concurred that distinguishing between the impacts associated with the elasticity of substitution and those resulting from non-neutral technical change may be difficult, particularly when the elasticity of substitution is different from 1 (see León-Ledesma, McAdam, & Willman, 2010, and references therein). However, assuming that the economy is characterised by a VES production function removes this fuzziness by presenting a clear cause and effect relationship between non-neutral/biased technical change and the elasticity of substitution, as the capital biased technical change emanating from a higher value of b leads to a rise in the elasticity of substitution in this instance.

The connection between capital deepening, the elasticity of substitution, and the share of capital created by the parameter b is of empirical relevance. For instance, Rodrik and Chen (1998) observes that East Asia experienced capital deepening and a simultaneous rise of the share of capital in total production over the decades of fast economic growth the region experienced, and points out that the growth literature often attributes these effects to either the higher elasticity of substitution between labour and capital or labour saving technical progress. However, such observations are typically made under the assumption that the underlying production function takes the CES form, which, as mentioned in the Introduction, cannot capture the endogeneity and variability of the elasticity of substitution highlighted in many studies. If, according to authors like Kazi (1980) and Bairam (1989), the VES form is indeed the correct functional specification of the production function, the contention of the CES literature—that capital deepening and a higher capital share are consequences of a higher elasticity of substitution—may be spurious and should therefore be viewed with caution. The alternative interpretation provided by the VES production function is that capital deepening, a higher share of capital in output, as well as a higher value of the elasticity of substitution are all the results of a higher value of the parameter b.

Having seen that b is an important technology parameter that determines the direction of technical change, the optimal capital-to-labour ratio and size of input shares, this discussion is incomplete without inquiring into what this parameter might represent. Essentially, it can be considered an institutional parameter that encapsulates a number of factors that might affect the direction of technical change in the economy. In the case of the CES production function, Mallick (2012b) provides a comprehensive treatment of the determinants of the elasticity of substitution between capital and labour, noting that "…the possible determinants of σ are technological progress, innovations, financial and other institutions, openness to trade, degree of unionisation, and the country's inclination towards socialist ideas." (p. 685). While the factors that that determine the value of σ in the CES production function are likely to be analogous to the characteristics the parameter b represents, the fact that the parameter b affects factor shares, factor intensity, and the direction of technical change warrants some further discussion on this matter.

C. I. Jones (2005) notes that a country can achieve a high elasticity of substitution only if it has access to the appropriate technology to support such a choice. Hence, a higher value of b primarily captures a technology that enables greater substitution between capital and labour. However, as a higher value of b leads to capital biased technical change, government policy relating to investment is likely to be an important determinant of the value of b. Government policy could sometimes distort the direction of investment and create an artificial bias towards one factor. For example, Yuhn (1991) points out that, in the 1970s, the South Korean government's tax policy led to massive investments in capital, which raised the capital per worker. In the context of the VES production function, such a change could be captured through a higher value of b. Typically, the ideals of the ruling party may affect the value of b and thereby the direction of technical change in the economy. Usually, conservative political parties might encourage capital-biased technical change by encouraging investments that raise the relative price of capital, while more liberal political parties that favour equity are unlikely to advocate a decline in the labour share and would thereby discourage capital-biased technical change. Moreover, Macpherson (1990) shows that the size and density of trade unions is positively related to the share of labour. Powerful trade unions could exert pressure on governments and the private sector to maintain the share of labour in production at an artificially high level, thereby preventing a decline in wages created by a falling share and marginal product of labour. In the context of the VES production function, this contributes towards a reduction of the value of b.

Having discussed the properties of the VES production function, and explored the role and importance of the parameter b in determining a number of key features of the production side of the economy, we apply the VES production function to a Diamond model in the next section.

3.3 THE VES PRODUCTION IN A DIAMOND OVERLAPPING GENERATIONS MODEL

We consider a standard two period overlapping generations model where the agent works during youth and spends her old age in retirement. Population in each period is normalised to 1. Time is discrete and is given by $t=0,1,2,\ldots$. The utility of an agent born at time t is given by:

$$U_t = u(c_t) + \beta u(c_{t+1}), \quad 0 < \beta < 1 \tag{3.3.1}$$

In (3.3.1) above, the consumer derives utility from youthful consumption c_t and old age consumption c_{t+1}. The parameter β is the discount factor.

In youth, the agent supplies one unit of labour and earns a wage w_t which she utilises on consumption c_t and savings s_t. The agent finances old age consumption c_{t+1} with her savings which accumulate a gross interest of R_{t+1}. She faces the following budget constraints in youth and old age respectively:

$$c_t + s_t = w_t \tag{3.3.2}$$

$$c_{t+1} = R_{t+1} s_t \tag{3.3.3}$$

The agent's utility maximisation problem yields the following standard FOC:

$$u'(c_t) = \beta R_{t+1} u'(c_{t+1}) \tag{3.3.4}$$

If we assume log utility such that $u(c_i) = \ln(c_i)$ for $i = t, t+1$; we get the following optimal solutions for consumption in both periods and savings:

$$c_t = \frac{w_t}{1+\beta} \tag{3.3.5}$$

$$c_{t+1} = \frac{\beta R_{t+1} w_t}{1+\beta} \tag{3.3.6}$$

$$s_t = \frac{\beta w_t}{1+\beta} \tag{3.3.7}$$

The intensive form production function for this economy is given by:

$$y_t = k_t^a (1 + abk_t)^{1-a} \tag{3.3.8}$$

During youth the agent supplies a single unit of labour within a competitive labour market and receives a wage w_t such that:

$$w_t = f(k_t) - k_t f'(k_t) = \frac{(1-a)y_t}{1+abk_t} \tag{3.3.9}$$

The gross rate of return on capital is given by:

$$R_{t+1} = 1 + f'(k_{t+1}) = 1 + \frac{ay_t}{k_t} + \frac{(1-a)aby_t}{1+abk_t} \tag{3.3.10}$$

Using the fact that $k_{t+1} = s_t$, we can now derive the evolution of the per capita capital stock for this economy which is given by:

$$k_{t+1} = \frac{\beta(1-a)}{(1+\beta)} \left(\frac{k_t}{1+abk_t} \right)^a \tag{3.3.11}$$

At steady state, $k_{t+1} = k_t = \bar{k}$. Therefore, the steady state capital stock is defined by the following implicit equation:

$$\bar{k}^{1-a} (1+ab\bar{k})^a - \frac{\beta(1-a)}{(1+\beta)} = 0 \tag{3.3.12}$$

The details of stability analysis carried out by linearising equation (3.3.11) and using equation (3.3.12) are provided in Appendix B. This analysis yields the following:

Proposition 1: *When $b \geq 0$, the steady state is always unique and stable.*

Our analysis in Appendix B shows that the steady state is unique and stable for $b > 0$. Nevertheless, since the VES production function takes the Cobb-Douglas form when $b = 0$, and it is well-known that the Cobb-Douglas always yields convergence towards a stable steady state in the Diamond model, we can conclude that as long as long as $b \geq 0$, an economy characterised by a VES production function always converges to a stable steady state.

When $b < 0$ there are three possible outcomes: there could be no non-trivial steady state, there may be a single stable steady state, or there could be two steady states, of which one is stable and the other is not. The number of non-trivial steady states depends on the value of a such that when $a > \frac{1}{2}$, one or two non-trivial steady states always exist. However, when $a < \frac{1}{2}$, either there is one non-trivial steady state, or it may be possible that there is no non-trivial steady state.

Proposition 1 is rather similar to the result obtained by Miyagiwa and Papageorgiou (2003), who demonstrate that in the context of a standard two period Diamond model with a normalised CES production function, for $\sigma > 1$, there is a unique steady state capital stock while for $\sigma < 1$ there can be two or no steady states. Like in Miyagiwa and Papageorgiou (2003), in the case of the VES form, a positive value of b, which results in $\sigma > 1$, is always associated with a unique and stable steady state.

The fact that the per capita capital stock has an upper bound of $\frac{1}{|b|}$ has important implications on the economy's long run growth trajectory in that it completely blocks the possibility of unbounded growth. Our analysis in Appendix B reveals that the parameter a, which was interpreted earlier as the pure share of capital in the production function, is a crucial determinant of the existence and the number of steady states. We are able to show that when $a > \frac{1}{2}$, the economy may have one or two steady states. If only one steady state exists, we demonstrate that it is a stable one. If two steady states exist, we demonstrate in Appendix B that only the lower value of the two is stable. Given the upper bound on the per capita capital stock, as demonstrated in Figure 3.3 below, if the economy starts off with an initial capital stock which is above the unstable steady state value, its capital stock will continue to grow until the upper bound is reached.

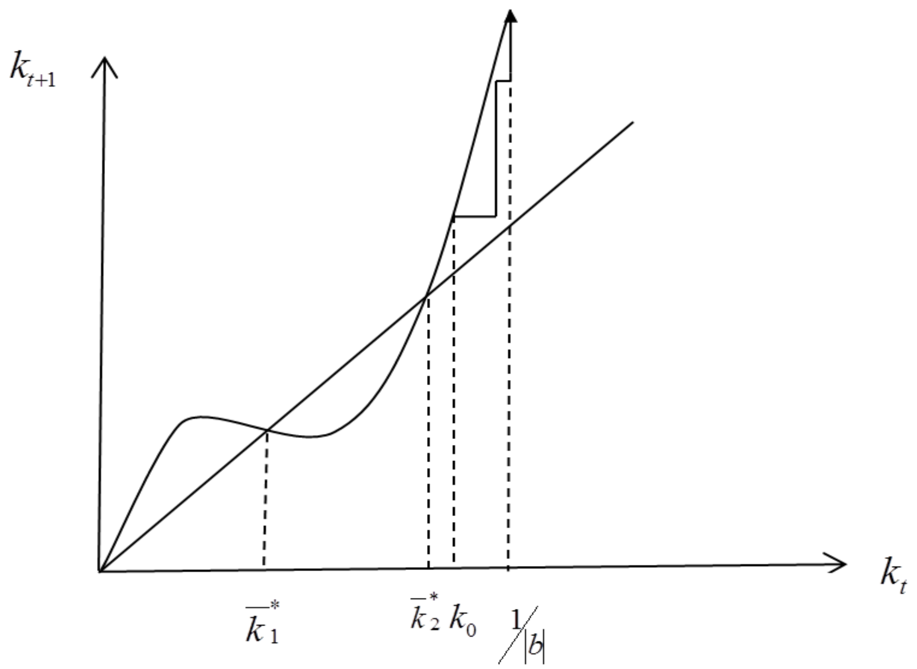


Figure 3.3: Growth trajectory when initial capital stock is above $\bar{k}_2^*$

On the other hand, when $a < 1/2$, the economy could either have a single stable steady state, or it may be characterised by no non-trivial steady state value, which takes a value below $1/|b|$. If a non-trivial steady state does not exist, it implies that the economy converges to the trivial steady state at $\bar{k} = 0$. In other words, the economy falls into a poverty trap in the long run, implying that a negative value of $b,$ when coupled with a value of a below $1/2$, could have a potentially detrimental impact on the economy.

Thus, in summary, a negative value of b presents three possible trajectories for the economy: it could fall into a poverty trap, converge towards a stable steady state, or achieve the upper bound of capital stock. The upper bound on the capital stock suggests that an economy to achieve unbounded growth. However, recall that the *AK* form is a limiting case of the VES production function that occurs when $a = 1$. As discussed, for instance, in Romer (1986, 1987) and Lucas (1988), when one adopts a broader interpretation of capital to include both human and physical capital, the AK form can yield unbounded growth in the Diamond model. Hence, the ability of an economy to achieve unbounded growth under the assumption of a VES production function crucially depends on the value of the parameter a

In relation to other related extant literature, our results are different from those of Klump and Preissler (2000), who demonstrate that when the CES production function is applied to a Solow-Swan model, the resulting steady state is always unique and stable. Furthermore, our results also differ from the findings of Karagiannis et al. (2005) who show that when the VES production function is applied to a standard Solow-Swan model, the economy could display unbounded endogenous growth for the case $b > 0$ if the condition $sA(ba)^{1-a} > n$ is satisfied, where s is the savings rate, A is an exogenous technological parameter and n is the rate of population growth. These authors note that when the elasticity of factor substitution is greater than 1, and hence, the marginal product of capital is bounded from below, it is possible for the economy to display unbounded endogenous growth. On the other hand, when $b < 0$ Karagiannis et al. (2005) show that the economy could either reach a unique steady state when $sA(ba)^{1-a} < n$ or reach a corner solution where $k = -1/b$ when $sA(ba)^{1-a} > n.$[19] A possible reason for the differences in the outcomes may emerge from the use of different frameworks. While in the Solow-Swan model, a fixed proportion of output per capita is saved, in the Diamond model with logarithmic preferences, the agent saves a fixed proportion of wage income. Although labour is the non-essential input in the VES production function, within an overlapping generations framework with logarithmic preferences, the accumulation of capital—which is the essential input—entirely depends on the wage received by the agent. Hence, these dynamics could introduce instability into the system.

Next we present another proposition, the proof of which is supplied in Appendix C:

Proposition 2: *For any non-trivial, locally stable steady state, the per capita capital stock $\bar{k}$ is falling in the value of the parameter b. However, for any feasible steady state, regardless of whether it is stable or not, steady state output $\bar{y}$ is rising in b.*

Proposition 2 shows that there is a pronounced benefit accruing to an economy due to a higher value of b as it enables the economy to achieve a higher output per worker at steady state with a lower capital stock. This result ties up with, and in fact accentuates, the capital biased technical change created by a higher value of b, which was discussed in detail in Section 3.2. While capital biased technical change would typically mean that the economy can produce more output with the same amount of capital as before, at steady state, the capital bias caused by a higher value of the

[19] An observation overlooked by Karagiannis et al. (2005) is that when $b < 0$, the expression $sA(ba)^{1-a}$ takes a complex value, and the economy may therefore be incapable of achieving a steady state at all.

parameter b actually allows the firm to produce more output using *less* capital per worker. Proposition 2 is similar in spirit to Klump and De La Grandville (2000), who demonstrate that under the assumption of a CES production function, a higher elasticity of substitution enables the economy to produce a higher level of output per capita. This result also supports the idea put forward by Kazi (1980), who argues that a low elasticity of factor substitution might be a reason for slow growth and development in many developing economies. However, as discussed in the Introduction, the elasticity of factor substitution can vary between industries belonging to a particular industry too. Given such heterogeneity within a particular economy, Kazi (1980) suggests that the economy's productivity would rise if investments are geared towards industries with a higher value of factor substitutability.

However, it is important to note that in Section 3.2, we demonstrated that a higher value of b results in a steeper long run expansion path, implying that the capital stock per capita increases. This can be referred to as the intensification effect. At the same time, it results in an inward shift of the isoquant, implying that an economy characterised by higher value b could produce a given level of output with less capital and labour, resulting in a productivity effect. At steady state, a higher value of the technology parameter b results in the economy producing a higher output with a lower steady state capital stock, implying that the productivity effect dominates over the intensification effect.

Another interesting observation in our study relates to intergenerational inequality. As shown in Section 3.2, in the VES production function, MP_K is increasing in b while MP_L is falling in it. In the presence of competitive factor markets, factor rewards are $r = MP_K$ and $w = MP_L$ respectively. In the context of a standard

Diamond model, where the young agents supply labour in return for a wage and the old agents are the owners of capital. Hence, the ratio $\gamma = {MP_K}/{MP_L}$, which was defined earlier as the capital bias, is also a measure of intergenerational income inequality. Furthermore, from equations (3.3.5) and (3.3.6), the ratio of the consumption of the old to that of the young is given by:

$$\frac{c_{t+1}}{c_t} = \beta R_{t+1} = \beta[1 + f'(k_{t+1})] \qquad (3.3.13)$$

From Lemma 1, we know that $f'(k_{t+1})$ is increasing in b, implying that a higher value of b exacerbates intergenerational consumption inequality too. Hence we have:

Proposition 3: *Intergenerational income and consumption inequality are rising in the value of b.*

Therefore, in this economy, the higher output per capita associated with a higher value of b comes at the cost of higher inequality. In fact, Proposition 3 provides a potential explanation for the possible divergence in inequality created by increases in the returns to capital suggested by Piketty and Goldhammer (2014). In the context of our model, the capital biased technical change created by a higher value of b leads to a redistribution of relative factor rewards in favour of the owners of capital, thereby exacerbating inequality.

Delving further into the mechanism behind this outcome, in the case of the CES production function, Irmen and Klump (2009) distinguish between three effects arising due to a change in the elasticity of substitution: the efficiency effect, distribution effect and acceleration effect. In the context of the normalised CES production function, the authors define the efficiency effect as the effect of the elasticity of substitution on output, i.e. $\frac{\partial y}{\partial \sigma}$. The distribution effect is defined as the effect of the elasticity of substitution on the share of capital in production i.e. $\frac{\partial s_k}{\partial \sigma}$. The efficiency and distribution effects combine to create the acceleration effect, which is the way in which the elasticity of substitution affects the speed of capital accumulation in the economy.

In the case of the VES form too, similar effects are present, but they can be analysed in a more straightforward manner to manner relative to b. The efficiency effect in the VES form occurs when a higher value of b directly results in a higher output for any given per capita capital stock and, as demonstrated in Proposition 2, it is also associated with a higher output per capita at steady state, i.e. $\frac{\partial y}{\partial \sigma} > 0$ and $\frac{\partial \bar{y}}{\partial \sigma} > 0$. The distribution effect occurs because a higher value of b, through its impact on factor rewards and shares, causes a redistribution of factor incomes from labour to capital. The positive efficiency and distribution effects lead to a clear positive acceleration effect i.e. a higher value of b improves the rate of capital accumulation in the economy. In contrast, in the case of the Diamond model with CES technology, as pointed out by Irmen and Klump (2009) with reference to the discrete time version Miyagiwa and Papageorgiou (2003) and the continuous time version by Irmen (2003), there is no guarantee of a positive acceleration effect. Even though the efficiency effect is always positive in the Diamond model, the decrease in the labour share resulting from a higher elasticity of factor substitution leads to a negative distribution effect. It may be possible that the negative distribution effect dominates the positive distribution effect, leading to a negative acceleration effect.

3.4 CONCLUSION

In this paper, we incorporate a VES production function into an otherwise standard Diamond model. We explore how the value of an exogenous parameter, labelled b, which affects the direction of technical change in the economy and is a determinant of the elasticity of substitution between labour and capital, affects the dynamics of growth. We observe that for positive values of the parameter b, which are associated with an elasticity of substitution between labour and capital that is greater than 1, the economy can always reach a unique and stable and steady state, which is feasible with the condition that the associated elasticity of substitution is positive. As such, when the value of b is positive, the behaviour of the economy, both during transition and at steady state, is identical to that associated with the standard Solow growth model. On the other hand, for negative values of b, for which the elasticity of substitution is between 0 and 1, we show that the economy could either fall into a poverty trap where the capital stock per worker is zero, or there could be two steady states of which one is stable and the other is not. For negative values of b, there exists certain upper bound for the per capita capital stock, which implies that, if the economy's initial capital stock is above the unstable steady state value, the capital stock would continue to grow until it reaches this upper bound and remain fixed at this level thereafter.

We also observe that a higher elasticity of factor substitution enables the economy to produce a higher level of output at steady state with a lower per capita capital stock. These results lead to the conclusion that a positive value of the technology parameter b yields multiple benefits upon the economy by ensuring a stable steady state and enabling it to achieve higher productivity. However, a higher value of the technology parameter b entails an efficiency-equity trade-off, as intergenerational inequality is rising in the value of this parameter. Hence, whether a higher value of the technology parameter b can indeed be perceived as desirable depends on how one weighs the benefits accruing from a high elasticity of substitution in the form of efficiency and stability against the loss inflicted upon the economy due to the exacerbation of inequality.

A potential direction for future research would be obtaining cross-country estimates for the value of b using non-linear regression techniques. Such estimates for individual countries could provide new cross-country estimates for the extent of capital-biased technical change experienced by countries, as well as estimates for the values of the elasticity of substitution for different countries, and supply rich insights into the heterogeneity in factor shares and rewards observed in practice. Some of the potential theoretical offshoots of this study include considering an alternative functional form of the VES production function in which labour is the essential input, and exploring the possibility of the political economy determination of an "optimal" value of b.

Chapter 4: Public and Private Education Expenditures, Variable Elasticity of Substitution and Economic Growth

4.1 INTRODUCTION

Given conventional wisdom regarding the positive externalities associated with education, its provision by the state is often regarded as desirable. In addition to the state, parents also play a critical role in educating their children. According to Article 26 of the 1948 Universal Declaration of Human Rights "parents have a prior right to choose the kind of education that shall be given to their children" (T. James, 2004). Generally, most decisions about a child's education, at least until the tertiary level, are made by parents.[20] The importance of parental expenditure in financing children's education is also highlighted by Becker (2009, pp.367-369) and Bräuninger and Vidal (2000), who note that in the presence of borrowing constraints, children cannot resort to the market to finance their education, and hence, have to rely on their parents for funding. This means that parental altruism is a key determinant of private education expenditure. Such private spending on education can exacerbate inequality, as more affluent parents can typically incur greater private expenditure on their children's education. Hence, both public and parental educational expenses can have important

[20]At the tertiary level, in some contexts, the presence of higher education income contingent loans, an arrangement that originated in Australia and was adopted by many other countries successfully over time (Di Gropello, 2012), enables a child to exercise considerable if not sole decision-making power relating to her educational choices.

implications on the accumulation of human capital in an economy. Therefore, in addition to considering the role of state provision, Glomm (1997) and Das (2007) note that a modelling construct exploring the macroeconomic impacts of education spending should ideally take into account the role played by parents in financing their children's education.

Studying public and private expenditures in conjunction with one another naturally brings into play another dimension of relevance to this issue: namely the degree to which agents view these expenditures as substitutable or complementary to each other. Much of the extant theoretical literature exploring the public-private mix of education expenditures and its impact on economic growth and political economy outcomes typically assumes that the two types of expenditure are either gross substitutes or complements, and sparse attention has been paid to the issue of how *variations* in the degree of substitutability between the two types of expenditure can affect macroeconomic outcomes. For instance, a number of studies have explored the public versus private education divide (see Glomm & Ravikumar, 1992; Epple & Romano, 1996a; de la Croix & Doepke, 2004; Goldhaber, 1999, among others). These studies perceive publicly and privately funded education to be alternatives/substitutes to one another. On the other hand, private supplementation of publicly provided goods like education and health care, where public and private expenditures are regarded as complements, has been the focus of studies such as Epple and Romano (1996b) and Gouveia (1997). Another dimension of complementarity between public and privately funded education can emerge from an arrangement where publicly provided primary and secondary education is a prerequisite for undertaking tertiary education, which is privately financed, either by parents or through loans that students are liable to repay once they obtain employment. This arrangement has been explored in studies such as

Kaganovich and Zilcha (1999), Blankenau, Cassou, and Ingram (2007), and Arcalean and Schiopu (2010).

A possible reason for the paucity of research on the macroeconomic impacts associated with the degree of substitutability between public and private education inputs is the lack of empirical evidence relating to this issue (Bearse et al., 2005). Nevertheless, Bearse et al. (2005) make a notable contribution to the theoretical literature on this subject by explicitly exploring the political economy ramifications associated with different values of the elasticity of substitution between public and private education inputs. These authors incorporate publicly and privately provided educational services into a constant elasticity of substitution (CES) education production function. For the special case of perfect substitutes, their model reveals that parents do not enrol their children in private schools at all. In the general CES case, they demonstrate that a higher elasticity of substitution between public and private education expenditures results in everybody selecting public schooling. However, as the relative efficiency of the private sector declines, although public school enrolment diminishes, agents vote for a higher tax rate to fund public education, resulting in higher public education expenditure per student.

While Bearse et al. (2005) provide a number of interesting insights into the issue of substitutability between public and private education expenditures, especially the political economy outcomes, they do not specifically explore the associated long run macroeconomic impacts. Therefore, in this paper, our aim is to explore how the degree of substitutability between public and private education can affect the dynamics of human and physical accumulation, as well as optimal policy in an economy in the long run. To this end, we consider a three period overlapping generations model in which an individual's adult age human capital is determined by the education she receives in childhood and her parent's human capital. Output in this economy is produced using a standard, constant returns to scale, Cobb-Douglas aggregate production function in which physical and human capital are the inputs. A key point of distinction in our model is that education is "produced" using a variable elasticity of substitution (VES) education production function in which the inputs are public and private expenditures. The VES specification, originally discussed in Sato and Hoffman (1968) and Revankar (1971), when applied to the context of education, provides a tractable functional form, in addition to having some appealing properties of relevance in the context of education.

As will be discussed shortly, the VES construct allows exploration of how institutional and cultural factors can influence the elasticity of substitution between public and private education expenditures experienced by an agent, which in turn has an effect upon the specific combination of public and private education expenditures she chooses, and consequently, the economy's performance in the long run. This means that the VES form contains an important channel for determining long run macroeconomic outcomes that is not present in the CES case.

Clarifying this point, in the CES production function, the degree of substitutability between two inputs, as measured by the elasticity of substitution, remains static across different combinations of the two inputs and over time. While this feature may be reasonable in the context of a standard production function with capital and labour as the inputs, in the case of an education production function, the elasticity of substitution between public and private expenditures can be conditioned by the extent to which people *perceive* these two types of expenditures to be substitutable for one another, as well as the *mix* of public to private expenditures. In order to gain a deeper insight into these considerations, take the case of South Korea, where the cultural emphasis on educational attainment and the intense competition for admission to top universities has led to the development of a massive private tutoring industry in the country (Kim & J. H. Lee, 2010). Institutional and cultural factors have therefore resulted in South Korean parents perceiving supplementary tutoring to be a vital complement to publicly provided school education, which leads to a lower elasticity of substitution between public and private education inputs. However, in recent years, policy reforms, such as improving the quality of public schooling, giving individual schools greater autonomy, and setting national standards for public exams, have been implemented in South Korea in an effort to reduce the demand for private tutoring (C. J. Lee, H. Lee, & Jang, 2010). Such policy reforms have the potential to reduce parental expenditure on supplementary tutoring by impacting on parental perceptions about the extent of substitutability between public and private education, and in the context of the VES production function, such changes can alter the mix of public and private expenditures chosen by them.

In the VES form, the institutional and cultural factors that influence the representative agent's ability to substitute between public and private education expenditures can be captured by a parameter labelled b, which we refer to as the "aggregate substitutability parameter", as the elasticity of substitution between public and private expenditures is positively related to this parameter. Therefore, in our interpretation, it captures the relative uniformity of public education in comparison to private education, with a higher value of b being associated with a greater degree of uniformity between public and private education expenditures. This uniformity is intrinsically related to the quality of both inputs, given that they are only perceived as "uniform" if they are of comparative quality. The parameter is therefore an indicator of the quality of the education system.

In practice, a higher degree of uniformity between public and private education is likely to be observed in developed countries where both the public and private education sectors are likely to be of a high quality. Notice that in Figure 4.1 below, the least developed countries have a higher percentage of private enrolments at the secondary level. Relating this observation to our notion of aggregate substitutability, a potential explanation for this could be the possibility that the quality of the public education system in these countries is of a relatively poor quality, which encourages parents to enrol their children in private schools. Although enrolling children in private schools results in those parents opting out of the public system, we wish to emphasise to the reader that in the present exercise, public and private education are not considered to be completely independent alternatives for one another. Rather, we abstract from the two extremes of perfect substitutes and perfect complements in the context of the public and private education expenditures and use the VES form to pay attention to a range of intermediate values of the elasticity of substitution that fall between these two extremes. In almost every modern society, parents can choose between sending their children to a public school or private school. Furthermore, they can choose how much to spend on supplementing their children's school education with privately provided education services. The aggregate substitutability parameter, which is at the heart of our model, can be interpreted as a weighted average of all these individual decisions, which in turn are influenced by many cultural and institutional parameters. In our model, the representative agent takes this aggregate substitutability parameter as given when she maximises her utility.

At this point, it is worthwhile recalling another characteristic of the VES form discussed in Chapter 3 previously, which is that one input is "essential" for production. In other words, whatever the degree of aggregate substitutability may be, no output can be produced without this essential input. In contrast, the CES production function admits a similar feature only in the special case when the two inputs are perfect substitutes, in which instance the cheaper input is used for production. In the context of the present model, we view public education expenditure as the essential input.[21] . Hence, in our model, as mentioned in the preceding paragraph, the agent cannot opt out of public education. The choice of public education as the essential input is driven by the fact that in most countries and regions, public sector involvement in education is relatively higher than private sector involvement.[22] For instance, data on private school enrolment at the secondary level published by the UNESCO Institute of Statistics shows that in many countries, private enrolments account for a very small percentage of the total, suggesting that assuming public education to be the essential input is of empirical relevance. Further evidence to support this view is provided in Figure 4.1, which shows that despite considerable variations in the proportion of private school enrolment at the secondary level in different parts of the world, private enrolments are always below 50% of the total enrolments, implying that public education may indeed be the relatively more important input.[23]

[21] The motivation underlying this choice is discussed shortly. We also considered the alternative, somewhat counter-intuitive and counterfactual paradigm in which private education is the essential input. This analysis yielded some intuitively unappealing results in addition to indeterminacies with respect to certain ranges of parameter values. The results of this analysis are available upon request.

[22] There are exceptions however. In the UK, private school enrolment at the secondary level was 66% in 2013, despite the government spending on secondary education accounting for around 2.30% of GDP (World Bank, 2016). A possible explanation for this high private enrolment may be that a significant part of the public funding goes towards subsidising private education, which provides further motivation to our view that public expenditure on education is likely to be the essential input even in some countries where private enrolment is very high.

[23] According to our interpretation, private education expenditure is incurred by parents on education services provided by third parties during a child's formal school years. It does not involve the time and opportunity costs involved with parents educating their children. Private education provided by parents at home, such as teaching a child to walk and talk and guiding them to adopt good manners is an essential part of a child's upbringing, especially during the formative years of a child's life. However, this type of education is not considered in our model.

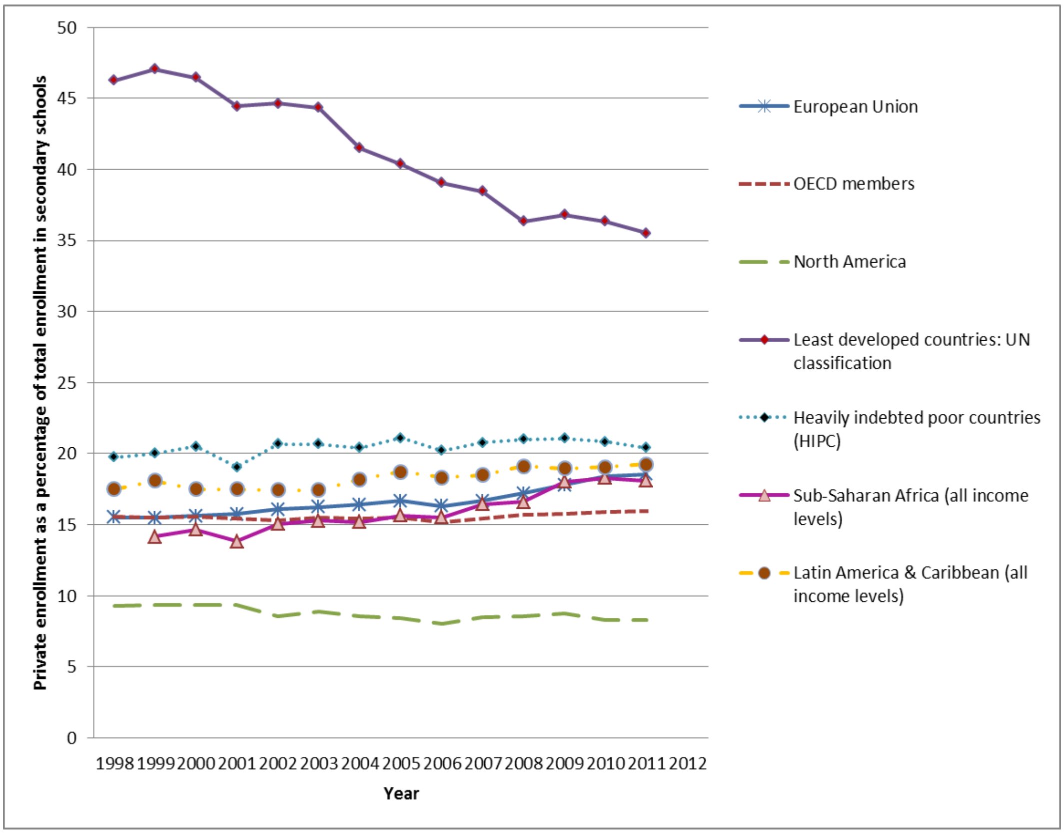


Figure 4.1 Private enrolment as a percentage of total secondary school enrolment

Source: World Bank data

As regards the key results of our paper, we are able to analytically prove that higher aggregate substitutability between public education expenditures, as captured by a higher value of the parameter b, yields better long run outcomes, in terms of higher steady state stocks of human and physical capital per capita, and also enables the economy to transition towards the steady state faster. In our model, we restrict the range of values for the parameter b to ensure interior solutions for the agent's decision problem and that the elasticity of substitution between public and private education expenditures is always positive. Within this restricted range, we conduct stability analysis of the two dimensional discrete dynamical system formed by the human and physical capital accumulation functions. This stability analysis reveals that the economy can display monotonic or oscillatory convergence towards a stable steady state or display oscillatory divergence and subsequently display cyclic behaviour.

Given that in our model, all tax revenue goes towards the provision of public education, another interesting consideration is the manner in which the degree of substitutability between public and private education can affect political economy outcomes. To this end, we examine how the optimal tax rate is affected by the aggregate substitutability parameter b. We are able to analytically demonstrate that the optimal tax rate is rising in this parameter, implying the presence of greater support for tax financed education in economies where agents perceive public and private education expenditures to be closely substitutable for one another. This outcome is quite similar in spirit to the higher per capita public education expenditure that results when the relative efficiency of the private sector diminishes in Bearse et al. (2005).

As the aggregate substitutability parameter yields higher stocks of human and physical capital in this economy and also enables faster transitioning towards steady state, this suggests that an economy is likely to benefit from policies aimed at improving the degree of aggregate substitutability between public and private education expenditures. In countries where private education is regarded to be of a higher quality than public education, providing better public education improves the comparability between the two types of expenditures, which improves the range of educational choices available to individuals and reduces the need for private expenditures.

The rest of this essay is organised as follows: Section 4.2 presents the model, Section 4.3 looks at steady states and stability, Section 4.4 considers optimal policy, and Section 4.5 concludes the chapter. Several proofs and derivations are supplied in the Appendices.

4.2 THE MODEL

We consider a three period overlapping generations model where an agent spends her childhood studying, works in adulthood, and spends her old age in retirement. In adult age, the agent gives birth to a single offspring. The population is normalised to unity in each period. Time is discrete and is given by $t=0,1,2,\ldots$ At time $t=0$, each adult is endowed with h_0 units of human capital and k_o units of physical capital. There is a representative firm producing a single good, and a government that raises revenue for the purpose of financing public education.

4.2.1 Human capital and education

Public education expenditure e_t^G and private expenditure on education e_t^P combine to form a child's stock of education e_t. The "education production function" takes a variable elasticity of substitution (VES) form as follows:

$$e_t = \left(e_t^G\right)^a \left(e_t^P + abe_t^G\right)^{1-a} \tag{4.2.1}$$

The following standard parameter restrictions as discussed in Revankar (1971) hold: $0 \le a \le 1$, and $\frac{e_t^G}{e_t^P} < \frac{1}{|b|}$ when $b < 0$.[24]

In (4.2.1) above, as we mentioned previously, b is a catchall parameter that affects the aggregate degree of substitutability between public and private education expenditures. The parameter a can be interpreted as the "pure" share of public education in the education production function. We abstract from technological

[24] As will be shown, in addition to these standard restrictions, further restrictions need to be imposed on the value of b in our model to ensure interior solutions for decision variables.

progress in the education production function for simplicity and also assume that the education production function is characterised by constant returns to scale. It can be seen that when $e_t^G = 0$, $e_t = 0$. Hence e_t^G can be considered the essential input in the education production function.

The elasticity of substitution between private and public expenditures is:

$$\eta = 1 + b\left(\frac{e_t^G}{e_t^P}\right) \tag{4.2.2}$$

Note that in (4.2.2), if $b \geq 0$, $\eta \geq 1$ and if $-1 < b \leq 0$, $0 \leq \eta \leq 1$. We can see that the agent's elasticity of substitution between private and public education expenditures is positively and linearly related to the aggregate substitutability parameter b. Further, the elasticity of substitution is also impacted by the ratio of public to private education expenditures, with an increase in this ratio leading to a rise in the elasticity of substitution if b is positive, and a reduction in the elasticity of substitution if b is negative. As there is no empirical evidence regarding the link between the ratio of public-private expenditures and the degree of substitutability between them, the VES form allows the flexibility to explore both of these possibilities by varying the value of b.

The human capital production function is of the Cobb-Douglas form, with education acquired in childhood and parental human capital forming an adult's stock of human capital h_{t+1} as follows:

$$h_{t+1} = e_t^{\gamma} h_t^{\delta}, \qquad 0 < \gamma, \delta < 1; \ \gamma + \delta > 1 \tag{4.2.3}$$

In (4.2.3), parental human capital affects the child's human capital directly through the inclusion of parental human capital in the human capital formation equation, as well as indirectly through education, as higher human capital enables the parent to earn more, thereby making it possible for the parent to invest more in her child's education. We make the assumption of increasing returns to scale in the accumulation of human capital like in Glomm and Ravikumar (1998) and also abstract from time spent studying in order to improve analytical tractability.

4.2.2 Production

Production in this economy follows a Cobb Douglas functional form with human and physical capital as the inputs like in Boldrin (2005), and takes the following intensive form:

$$y_t = k_t^{\lambda} h_t^{1-\lambda} \tag{4.2.4}$$

The competitive interest rate and wage rate per unit of human capital are given by:

$$r_t = \lambda \left(\frac{k_t}{h_t} \right)^{\lambda - 1} \tag{4.2.5}$$

$$w_t = (1-\lambda) \left(\frac{k_t}{h_t} \right)^{\lambda} \tag{4.2.6}$$

From (4.2.6) above, an adult's earnings in period t is her wage per unit of human capital w_t times her stock of human capital h_t which is accumulated according to (4.2.3).

4.2.3 Government

The government imposes a proportional tax τ in each period on the adult agent's earnings to finance public education expenditure and runs a balanced budget. Hence,

$$e_t^G = \tau w_t h_t \tag{4.2.7}$$

4.2.4 The agent's problem

The agent's utility at time t is given by:

$$U = u(c_{t+1}) + \beta u(c_{t+2}) + \theta v(h_{t+2}) \ , \ 0 \le \beta \le 1, \ \ \theta > 0. \qquad (4.2.8)$$

The agent derives utility from consumption in adulthood c_{t+1} and consumption in old age c_{t+2}. In old age, the agent also derives utility from her child's human capital h_{t+2}. β is the discount factor. The parameter θ is the product of the discount factor β and another "warm glow" parameter, which measures the satisfaction a parent receives from her child's human capital.

This agent faces the following budget constraints in adulthood and old age respectively:

$$c_{t+1} + e_{t+1}^{P} + s_{t+1} = w_{t+1}h_{t+1}(1-\tau) \qquad (4.2.9)$$

$$c_{t+2} = R_{t+2}s_{t+1} \qquad (4.2.10)$$

According to (4.2.9) above, in adulthood, the agent inelastically supplies h_{t+1} units of human capital, and receives a wage of w_{t+1} per unit of human capital supplied. The government levies a proportional tax of τ on her labour income to finance public education. Hence, her after tax income is $w_{t+1}h_{t+1}(1-\tau)$. She utilises her after tax income on consumption c_{t+1}, private expenditure on her child's education e_{t+1}^{P} and savings s_{t+1}. According to (4.2.10), in old age, the agent uses her savings that have accumulated a gross real return of $R_{t+2} = 1 + r_{t+1}$ to finance her consumption c_{t+2}.

Thus, the agent's utility maximisation problem involves maximising (4.2.8) subject to (4.2.9) and (4.2.10). This yields the following FOCs:

$$U^{'}(c_{t+1}) = \beta R_{t+1}u^{'}(c_{t+2}) \qquad (4.2.11)$$

$$U^{'}\left(c_{t+1}\right)=\theta v^{'}\left(h_{t+2}\right) \tag{4.2.12}$$

(4.2.11) is the standard consumption Euler equation that captures the fact that the consumer cannot improve her utility by shifting consumption between periods when she maximises utility. (4.2.12) captures the idea that investing in her child's education causes the agent to forego utility from adult age consumption in order to enjoy a higher utility from her child's stock of human capital in old age. When the agent maximises utility, the marginal cost of a unit of private education spending in terms of the foregone adult age consumption should be equal to the marginal benefit the agent enjoys in terms of the additional human capital her child acquires.

4.2.5 Competitive equilibrium

A competitive equilibrium in this environment is a sequence of consumption, savings and private education expenditure $\left\{c_{t+1}, c_{t+2}, s_{t+1}, e_{t+1}^{P}\right\}_{t=0}^{t=\infty}$ chosen by the agent that satisfies the FOCs (4.2.11) and (4. 2.12), taking factor prices represented by (4.2.5) and (4.2.6) as given; a sequence of input and output choices $\left\{y_{t,} k_t, h_t\right\}_{t=0}^{t=\infty}$ made by the representative firm according to (4.2.4) so as to maximise profits, taking factor prices and government policy as given; a sequence of factor prices $\left\{r_t, w_t\right\}_{t=0}^{t=\infty}$ as given by (4.2.5) and (4.2.6), such that the markets for physical and human capital clear; and a sequence of government policies $\left\{e_t^{G}, \tau_t\right\}_{t=0}^{t=\infty}$ such that the government's budget as given by (4.2.7) in each period is balanced.

Assuming logarithmic preferences, equations (4.2.13) to (4.2.16) below are the optimal solutions to the agent's decision problem.

$$c_{t+1}=\frac{w_{t+1}h_{t+1}\left(1-\tau+ab\tau\right)}{1+\beta+\theta\gamma\left(1-a\right)} \tag{4.2.13}$$

$$e^{P}_{t+1} = w_{t+1}h_{t+1}\left[\frac{\theta\gamma(1-a)(1-\tau)-ab\tau(1+\beta)}{1+\beta+\theta\gamma(1-a)}\right] \tag{4.2.14}$$

$$s_{t+1} = \frac{\beta w_{t+1}h_{t+1}(1-\tau+ab\tau)}{1+\beta+\theta\gamma(1-a)} \tag{4.2.15}$$

$$c_{t+2} = \frac{\beta R_{t+2}w_{t+1}h_{t+1}(1-\tau+ab\tau)}{1+\beta+\theta\gamma(1-a)} \tag{4.2.16}$$

Let $m = -\frac{(1-\tau)}{a\tau}$ and $n = \frac{\theta\gamma(1-a)(1-\tau)}{a\tau(1+\beta)}$. Interiority of the above solutions occur only in the range $m < b < n.$ However, in any given period, the elasticity of substitution between public and private education inputs η, given by equation (4.2.2), should be bounded from below so that $\eta > 0$. Hence, the above range of values for b needs to be modified to incorporate this constraint. Given $p = \frac{-\theta\gamma(1-\tau)}{\tau[(1+\beta)+\theta\gamma]}$, the range of values of b necessary to ensure interiority *and* a positive elasticity of substitution between public and private inputs is: $p < b < n$. A detailed derivation of this range is provided in Appendix F. The analyses presented in the remainder of the paper assume that the parameter b conforms to this range.

From (4.2.14), we can see that as long as $b > \frac{-\theta\gamma(1-a)}{a(1+\beta)}$, public education expenditure crowds out private education spending. This is because a higher value of the aggregate substitutability parameter b means that there is a greater degree of similarity between private and public education inputs in terms of access and quality. Thus, as indicated by equations (4.2.13) to (4.2.16), a higher value of the aggregate substitutability parameter b causes the agent to undertake lower private investment in education, causing consumption in both periods and savings to rise. Hence, from (4.2.8), we have:

Proposition 1: *A higher value of b results in a higher utility for the agent.*

The intuition behind Proposition 1 lies in the idea that a higher value of the aggregate substitutability parameter b, which impacts the elasticity of substitution positively, implies that private education spending is a closer substitute for public spending. This reduces the incentive for a parent who already supports the public education system by paying a mandatory tax to undertake private expenditure. Typically, parents would be keen to undertake more private expenditure on education only if private education services can augment a child's learning experience because they are of a superior quality to public expenditure, implying a lower degree of substitutability between the two inputs. In such an instance, private education services play a critical role in enhancing and augmenting the child's stock of education, and parents would therefore be keen to spend on such services in order to compensate for the low quality of public education their children receive. Therefore, when aggregate substitutability is high, the lower private education expenditure that results enables agents to undertake higher consumption and savings, leading to a higher level of utility from consumption during both adulthood and old age, thereby resulting in a higher lifetime utility.

4.3 STEADY STATE AND STABILITY ANALYSIS

In this section, we first derive and explore the behaviour of the steady state values of per capita physical and human capital. Then we explore the dynamic properties of the model and the conditions under which the steady state is stable.

4.3.1 Steady state analysis

First, note that savings in period $t+1$ are used to build the stock of physical capital in period $t+2$. Hence, in general equilibrium,

$$s_{t+1} = k_{t+2} \tag{4.3.1}$$

Using equation (4.3.1), and setting $j = \dfrac{(1-\tau+ab\tau)}{1+\beta+\theta\gamma(1-a)}$ the physical capital accumulation equation is:

$$k_{t+2} = \phi^1(k_{t+1}, h_{t+1}) = \beta j k_{t+1}^{\lambda} h_{t+1}^{1-\lambda} \tag{4.3.2}$$

Similarly, using (4.2.3), the human capital accumulation function is:

$$h_{t+2} = \phi^2(k_{t+1}, h_{t+1}) = k_{t+1}^{\lambda\gamma} h_{t+1}^{\delta+(1-\lambda)\gamma} (1-\lambda)^{\gamma} \left\{\tau^{a} [\theta\gamma(1-a) j]^{1-a}\right\} \tag{4.3.3}$$

(4.3.2) and (4.3.3) constitute of a two-dimensional system of non-linear first order difference equations. A steady state equilibrium of this system is a pair of values $(\bar{k}, \bar{h})$ such that $k_{t+2} = k_{t+1} = \bar{k}$ and $h_{t+2} = h_{t+1} = \bar{h}$. The steady state stocks of physical and human capital are therefore given by:

$$\bar{h} = \left\{(1-\lambda)^{\gamma} \tau^{a\gamma} j^{\frac{\lambda\gamma+(1-a)(1-\lambda)\gamma}{1-\lambda}} \beta^{\frac{\lambda\gamma}{1-\lambda}} [\theta\gamma(1-a)]^{(1-a)\gamma}\right\}^{\frac{1-\lambda}{1-\lambda\gamma-(1-\lambda)[\delta-(1-\lambda)\gamma]}} \tag{4.3.4}$$

$$\bar{k} = (j\beta)^{\frac{1}{1-\lambda}} \left\{(1-\lambda)^{\gamma} \tau^{a\gamma} j^{\frac{\lambda\gamma+(1-a)(1-\lambda)\gamma}{(1-\lambda)}} \beta^{\frac{\lambda\gamma}{1-\lambda}} [\theta\gamma(1-a)]^{(1-a)\gamma}\right\}^{\frac{1}{1-\lambda\gamma-(1-\lambda)[\delta-(1-\lambda)\gamma]}} \tag{4.3.5}$$

Before exploring the stability of this non-linear system, we examine the manner in which the steady state stocks of human and physical capital are affected by the value of the parameter b. The following proposition captures how the value of the aggregate substitutability parameter b impacts on the steady state stocks of human and physical capital:

Proposition 2: *A higher value of b enables an economy to reach a higher steady state level of physical and human capital.*

Note that j is monotonically increasing in b. As it is evident from equations (4.3.4) and (4.3.5) that the steady state stocks of human and physical capital are monotonically rising in j, it follows that they are also increasing in b. In the case of the conventional CES production function typically embedded in growth models, a higher elasticity of substitution between labour and capital has been demonstrated theoretically and empirically as a means of achieving a higher output per capita (see, for instance, Irmen, 2011; Karagiannis et al., 2005; Klump & De La Grandville, 2000; Mallick, 2012a). Proposition 2 is a somewhat heuristic extension of this notion, in that a higher elasticity of substitution between public and private education expenditures impacts an individual's stock of human and physical capital positively, thereby yielding a higher output per capita. As is evident from the education production function given by equation (4.2.3), the agent's stock of education, and thereby human capital, is increasing in the aggregate substitutability parameter. In terms of physical capital, a higher elasticity of substitution means that it is easier for an agent to substitute between public and private education. This means that a given level of human capital can be achieved with lower private spending, thereby enabling agents to save more in adulthood as is evident from equation (4.2.15), which shows that savings are increasing in the aggregate substitutability parameter b, which yields a higher stock of physical capital in the next period. This implies that the economy can enjoy a bigger stock of physical capital at any given level of human capital, and this is made clear if we look at how the ratio of physical to human capital at steady state is affected by the value of the aggregate substitutability parameter b. Dividing (4.3.4) by (4.3.5) gives us the ratio of physical to human capital, which is given by:

$$\frac{\bar{k}}{\bar{h}} = \bar{h}^{\frac{\lambda}{1-\lambda}} \left(j\beta\right)^{\frac{1}{1-\lambda}} \tag{4.3.6}$$

As $\bar{h}$ is monotonically increasing in b, it is clear that $\frac{\bar{k}}{\bar{h}}$ is also increasing in the aggregate substitutability parameter b. Thus, higher aggregate substitutability between public and private education expenditures causes the stock of physical capital to rise relative to the stock of human capital in steady state. A possible explanation for this could be gleaned by considering the choice a parent has between investing in her child's education and saving to finance old age consumption. When the value of the parameter b is higher, the return to private education expenditures is lower, which encourages people to undertake more savings, which causes the stock of physical capital to rise relative to human capital.

4.3.2 Stability analysis

Now, we analyse the stability of the system characterised by (4.3.2) and (4.3.3) which is of the form $X_{t+1} = \phi(X_t)$ where $X = \begin{pmatrix} k \\ h \end{pmatrix}$. We use the methods for studying the stability of non-linear first order discrete dynamical systems described in Galor (2007). First, we lag the two dimensional system characterised by equations (4.3.2) and (4.3.3) by one period, and we then employ a Taylor series expansion to linearise them around the steady state, which enables us to rewrite the system in the following form:

$$k_{t+1} = \phi^1(\bar{k},\bar{h}) + \frac{\partial \phi^1(\bar{k},\bar{h})}{\partial k_t}(k_t - \bar{k}) + \frac{\partial \phi^1(\bar{k},\bar{h})}{\partial h_t}(h_t - \bar{h}) \tag{4.3.7}$$

$$h_{t+1} = \phi^2(\bar{k},\bar{h}) + \frac{\partial \phi^2(\bar{k},\bar{h})}{\partial k_t}(k_t - \bar{k}) + \frac{\partial \phi^2(\bar{k},\bar{h})}{\partial h_t}(h_t - \bar{h}) \tag{4.3.8}$$

Upon evaluating the partial derivatives at the steady state values and substituting them into (3.7) and (3.8), and also setting $l = (1-\lambda)^{\gamma}\left\{\tau^{a}\left[\theta\gamma(1-a)j\right]^{1-a}\right\}$, the linearised system in matrix form can be expressed as follows:

$$\begin{bmatrix} k_{t+1} \\ h_{t+1} \end{bmatrix} = \begin{bmatrix} \beta j\lambda\left(\dfrac{\bar{k}}{\bar{h}}\right)^{\lambda-1} & \beta j(1-\lambda)\left(\dfrac{\bar{k}}{\bar{h}}\right)^{\lambda} \\ l\lambda\gamma\bar{k}^{\lambda\gamma-1}\bar{h}^{\delta+(1-\lambda)\gamma} & l[\delta+(1-\lambda)\gamma]\bar{k}^{\lambda\gamma}\bar{h}^{\delta+(1-\lambda)\gamma-1} \end{bmatrix}\begin{bmatrix} k_t \\ h_t \end{bmatrix} + \begin{bmatrix} 0 \\ l(1-\delta-\gamma)\bar{k}^{\lambda\gamma}\bar{h}^{\delta+(1-\lambda)\gamma} \end{bmatrix}$$

(4.3.9)

We have now approximated our original non-linear two-dimensional discrete dynamical system with a new system which is linear around the steady state that conforms to the form $X_{t+1} = AX_t + B$, where A is the Jacobian matrix of ϕ(X) evaluated at the steady state $\bar{X}$. The eigenvalues of A can be used to obtain a characterisation of the system.[25] We capture the different characterisations of the steady state as follows:

Proposition 3: Let $s_1 = \beta j\lambda\left(\dfrac{\bar{k}}{\bar{h}}\right)^{\lambda-1} + l[\delta+(1-\lambda)\gamma]\bar{k}^{\lambda\gamma}\bar{h}^{\delta+(1-\lambda)\gamma-1}$,

And $s_2 = \beta j\lambda l\delta\bar{k}^{\lambda(1+\gamma)-1}\bar{h}^{\delta+\gamma-\lambda(1+\gamma)}$.

The eigenvalues μ_1 and μ_2 of the Jacobian matrix A are given by:

$$\mu_1 = \frac{s_1 + \sqrt{s_1^2 - 4s_2}}{2} \qquad (4.3.10)$$

$$\mu_2 = \frac{s_1 - \sqrt{s_1^2 - 4s_2}}{2} \qquad (4.3.11)$$

[25] These characterisations derived from the linearised system provide an analysis of the local stability conditions.

Given these eigenvalues, the possible characterisations of the steady state are as follows:

i. If $s_1^2 > 4s_2$, $\frac{s_1 + \sqrt{s_1^2 - 4s_2}}{2} < 1$, we have distinct, real eigenvalues $0 < \mu_2 < \mu_1 < 1$, and the steady state is asymptotically stable, and there is monotonic convergence to the steady state. A sufficient condition for the emergence of this outcome is $\sigma < (1-\lambda)\gamma$.

ii. If $s_1^2 = 4s_2$, we have repeated, real eigenvalues such that $0 < \mu_2 = \mu_1 < 1$, the steady state is asymptotically stable, and both human and physical capital converge to the steady state at an equal rate.

iii. If $s_1^2 < 4s_2$, μ_1 and μ_2 are distinct complex eigenvalues with a positive real part since $0 < \frac{s_1}{2} < 1$, and the steady state is a clockwise spiral sink if $|\mu_1| < 1$ and a clockwise spiral source if $|\mu_1| > 1$.

In Proposition 3 above, case (i) resembles a long run outcome akin to that which occurs in the Solow growth model. If the initial value of human or physical capital is below the steady state, its value would monotonically increase until its steady state value is reached, while an initial value below the steady state leads to a monotonic decrease in the value until steady state is reached. As μ_1 is the eigenvalue associated with physical capital per capita k_t, the eigenvalues $0 < \mu_2 < \mu_1 < 1$ imply that physical capital per capita converges to its steady state value at a relatively faster rate than human capital per capita. However, once convergence to the steady state has occurred, the growth rate of physical and human capital per capita, as well as output per capita is 0, which is the outcome we observe in the standard Solow-Swan model. The phase diagram associated with this steady state is illustrated in Figure 4.2. Notice that in Figure 4.2, all the arrows point towards the steady state $(\bar{k}, \bar{h})$. The dotted straight lines indicate that if the initial stock of physical or human capital is equal to the steady state value, the value of the other variable whose initial value is above or below the steady state value will monotonically converge towards the steady state. The curved solid lines show the trajectory of the equilibrium path when the initial stocks of both human and physical capital are above or below the steady state. The small arrows on the sides, which point towards the dotted and solid lines, indicate that the two equilibrium trajectories can take only one of these two forms depending on the initial stocks of human and physical capital.

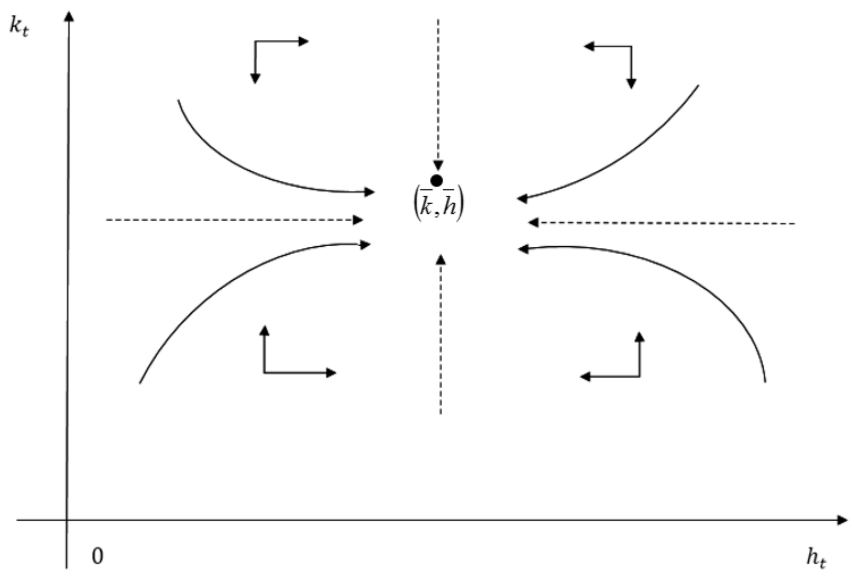


Figure 4.2: Phase diagram for the stable steady state

The steady state associated with the second case outlined in Proposition 3 is commonly referred to as a "focus." In this instance, both human and physical capital converge linearly to the steady state. This convergence occurs at an identical rate for both variables. Hence, the only difference between cases (i) and (ii) is the speed of convergence of the two variables to their steady state values. The phase diagram in this instance is depicted in Figure 4.3 below. In this instance, regardless of the initial stocks of human and physical capital, the family of straight lines surrounding the steady state combination of human and physical capital $(\bar{k}, \bar{h})$ indicates that there is systematic convergence to the steady state.

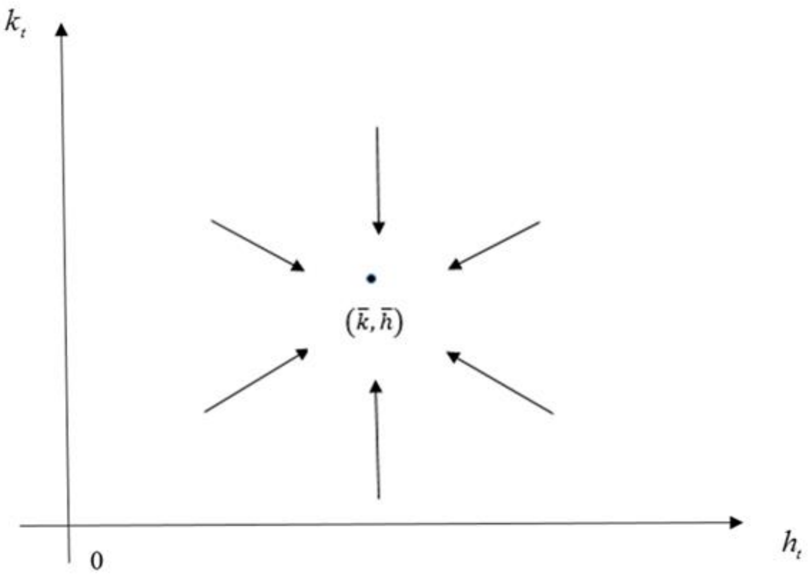


Figure 4.3: Phase diagram for the focus

When the eigenvalues are complex, as stated in case (iii) of Proposition 3, the two possibilities that emerge are either a spiral source or a spiral sink. In the case of a spiral sink, both physical and human capital per capita are characterised by oscillatory convergence to the steady state. These oscillations create fluctuations in the stocks of human and physical capital over time as the economy converges to the steady state. However, the steady state is asymptotically stable in this instance, implying that as $t \to \infty$, $k_t \to \bar{k}$ and $h_t \to \bar{h}$. We know that the spiral convergence to the steady state in this instance occurs in clockwise motions as the real part $\frac{s_1}{2}$ of both eigenvalues is positive. The phase diagram depicting the manner in which human and physical capital converge to the steady state is given in Figure 4.4. As we can see, regardless of whether the initial stocks of human and physical capital are higher or lower than the steady state values, the arrows that point inwards indicate that they converge over time to the steady state. Therefore, this outcome is also quite similar to cases (i) and (ii). The only difference in this instance is the non-monotonicity of the convergence path. The converging spiralling motions indicate that although both the stocks of human and physical capital are characterised by oscillations, the size of these oscillations grow smaller over time, and decay off once the steady state is reached.

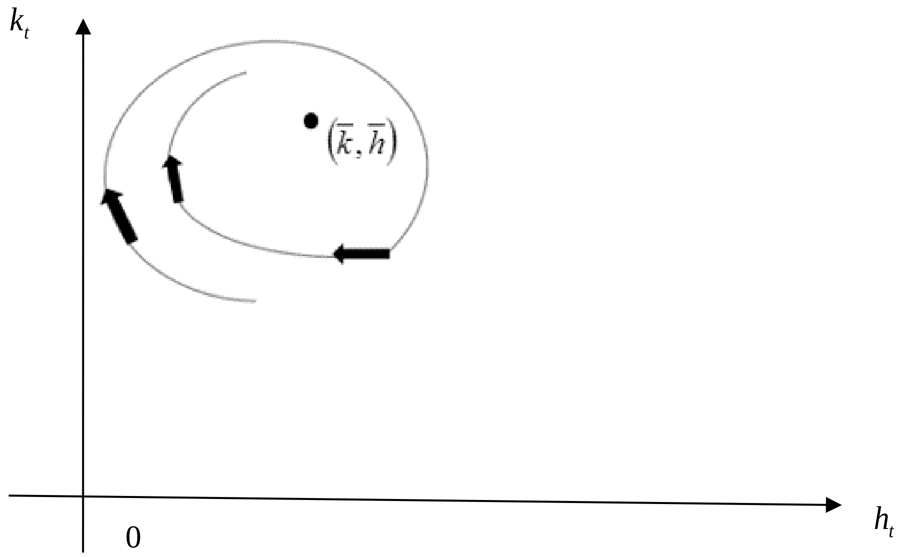


Figure 4.4: Phase diagram for the standard spiral sink

However, there is an aspect to the transitional dynamics that warrants further discussion. Recall that our model comprises of restrictions on the parameter b; as established earlier, for the elasticity of substitution between public and private education inputs to be positive and for interiority of the solutions to the agent's lifetime utility maximisation problem, the range $p < b < n$ should hold. This range for b also implies that there is a corresponding range of values for the stocks of human and physical capital in any given period given by equations (4.3.2) and (4.3.3). This implies that in any given period, both forms of capital should conform to certain ranges of positive values $k_{min} < k_t < k_{max}$ and $h_{min} < h_t < h_{max}$ as determined by these bounds for b.

While these bounds on human and physical capital are present across all the cases outlined in Proposition 3, they do not impact the convergence paths in the case of the *stable steady state* and the *focus*. The initial stock of human and physical capital should conform to these ranges, and the presence of these upper and lower bounds do not affect the monotonic convergence towards the steady state that occurs in these instances. However, with the *spiral sink*, as well as with the *spiral source*, which is considered shortly, these bounds create transitional dynamics that differ from the standard cases.

With the spiral sink, the phase portrait given in Figure 4.5 shows how the upper and the lower bounds affect the economy's convergence path. At point A, the standard convergence path falls outside the maximum human capital stock. Therefore, the economy moves along the upper bound until point B, where it meets the usual transitional path once again. Similarly, between points C and D, the standard path results in human capital stocks below h_{min}. Hence, the economy moves vertically along the bound given by h_{min} until it meets the usual convergence path at point D.

Nevertheless, the changes to the transition path that occur due to the presence of the upper and lower bounds do not prevent the economy from converging to the steady state in the long run.

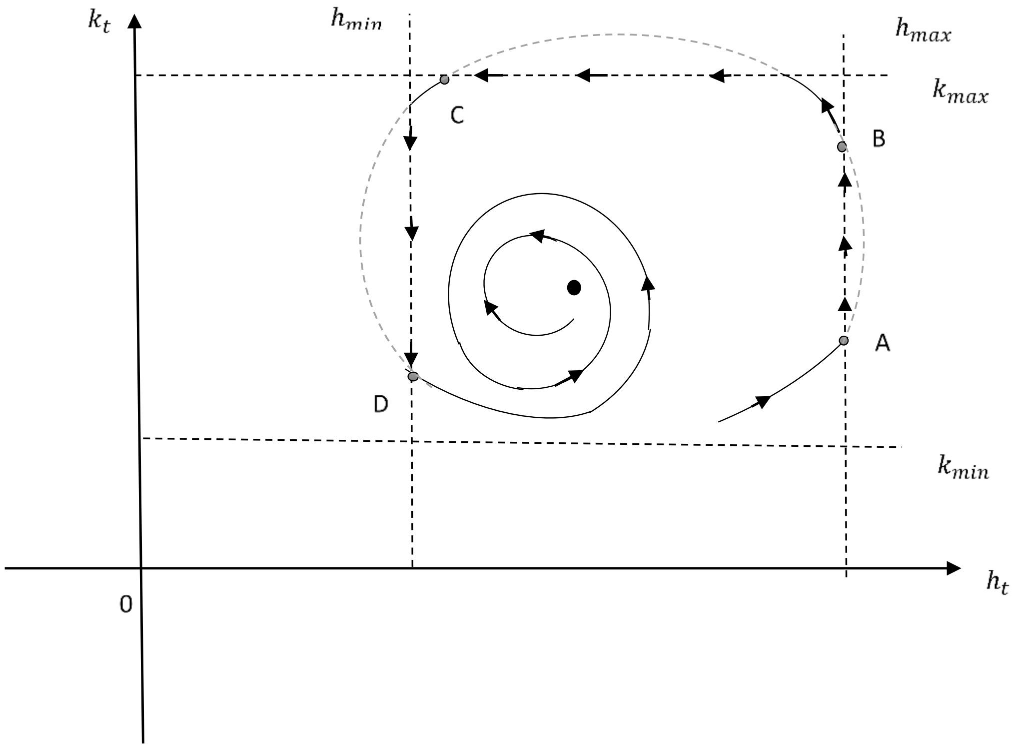


Figure 4.5: Modified phase diagram for the spiral sink

Now we consider the final possibility, which is the spiral source. In the case of the standard spiral source, the capital and human capital per capita display oscillatory divergence from the steady state. In this case too, we can infer that these oscillations occur in a clockwise motion, as both eigenvalues have a positive real part. The standard phase portrait in this instance is given in Figure 4.6. The outward-pointing arrows indicate that the oscillations grow bigger over time, implying that the stocks of human and physical capital keep diverging from their steady state values.

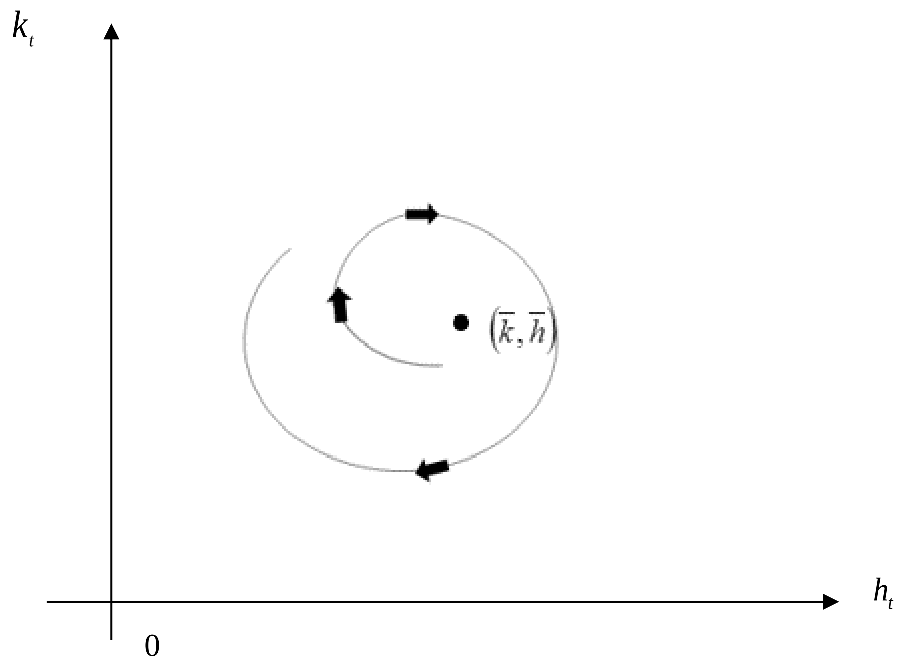


Figure 4.6: Phase diagram for the standard spiral source

In the case of the spiral source, it is possible that the large downward swings created by the oscillations may lead to negative values of the variables. In an economic context such as ours, however, negative values cannot occur for either of the variables in our two dimensional discrete dynamical system, or the other variables associated with consumption and production in the economy. Hence, in a context like ours, and indeed with most dynamical systems in economics, there should be restrictions preventing the occurrence of negative values. As a result of this boundedness of both human and physical capital from above and below, the occurrence of negative values in our system is avoided. However, the economy now follows a transition path that differs from the standard phase portrait for the spiral source. The modified phase portrait that takes into account the boundedness of the variables is given in Figure 4.7.

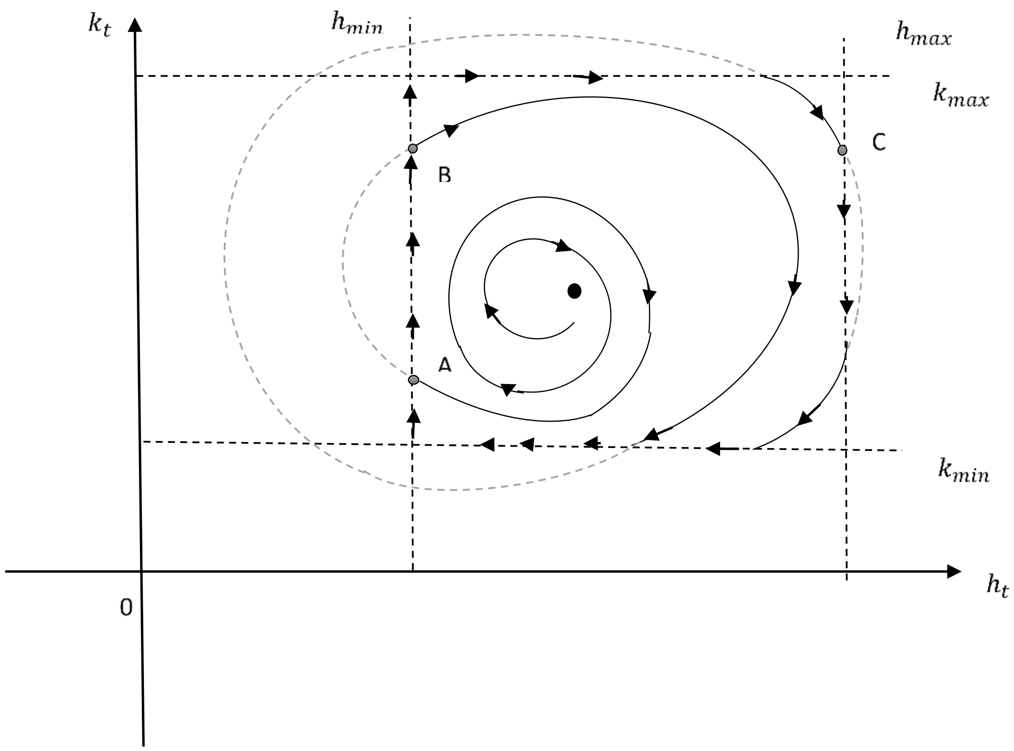


Figure 4.7: Modified phase diagram for the spiral source

In Figure 4.7, observe that the standard oscillations that take place in the case of a spiral source occur till point A, when the minimum value of human capital is reached. Beyond A, the economy cannot move along the usual dynamical path, which is depicted by the broken grey line. Instead, human capital remains at its minimum value, but since physical capital is still growing, there is still an upward movement along the line bounded by this minimum value, as given by the arrows. This movement enables the economy to connect to the usual path at point B. Similarly, at point C, human capital reaches its maximum value. Hence, as denoted by the downward arrows, while the stock of human capital remains at the maximum level, the motion of the physical capital stock still takes place until a connection with the oscillatory path occurs again. Once the diverging oscillations fall outside the bounds, the system can no longer connect to the spiralling movements along the dynamical path. The system will then move along the edges of the boundaries. This results in the system displaying periodic cycles in the long run along the path $(k_{min}, h_{min}) \rightarrow (k_{max}, h_{min}) \rightarrow (k_{max}, h_{max}) \rightarrow (k_{min}, h_{max})$. [26]

A number of studies have demonstrated the possible emergence of endogenous cycles in overlapping generations models. For instance, Michel and Venditti (1997) show that endogenous cycles may emerge in overlapping generations models with non-separable utility functions, and also in those characterised by bequests and a discount factor that is close to 1. Another study by Kitagawa and Shibata (2005) reveals that in an overlapping generations models with an AK type technology, lags in investment gestation could lead to endogenous growth cycles. Nevertheless, the emergence of regular cycles *after* a period of oscillatory divergence is an interesting outcome of our analysis, and is one that is not typically observed in the growth literature. In our model, the regular cycles that emerge in the long run are the result of the bounds on the parameters created by the functional form of the education production function and the related need to restrict the elasticity of substitution between public and private education expenditures to values greater than or equal to 0, as well as the need for interiority of solutions to the agent's utility maximisation problem. Nevertheless, the emergence of permanent cycles in the long run creates

[26] A number of numerical experiments that were carried out to glean insights into transitional dynamics led to monotonic convergence towards a stable steady state. However, it is important to note that the inability to find such combinations does not negate the existence of oscillatory convergence and divergence since the analytical results clearly show that these cases are a theoretical possibility.

uncertainty in the economy and the spiral source is hence less desirable in comparison to the other outcomes.

The variety of transitional dynamics revealed in our analysis shows that in the presence of concurrent public and private expenditures on education, as modelled by the VES form, monotonic convergence towards the steady state as observed in the standard neoclassical growth model may not always occur. The possibilities for oscillatory convergence and the emergence of permanent cycles after a period of oscillatory divergence demonstrates that our model is able to display some rich dynamics that are not frequently observed in the growth literature concerning investments in human capital.

It is also interesting to observe that the eigenvalues are monotonically increasing in the parameter b in the instances that monotonic convergences towards a stable steady state occurs, implying that a higher degree of aggregate substitutability between public and private education inputs enables an economy to converge to the steady state faster. This observation suggests that a higher degree of aggregate substitutability benefits an economy in two main ways. This observation is also consistent with extant literature that suggests that in the context of a standard production function with capital and labour, a higher degree of factor substitutability enables an economy to transition to the steady state faster (see Irmen, 2011 for a detailed discussion).

4.4 OPTIMAL POLICY

In this section, we look for the optimal value of τ a benevolent social planner may select in order to enable the representative agent to attain maximum utility. This is done by substituting the utility maximising values of c_{t+1}, e^{P}_{t+1} and c_{t+2} given by (4.2.13), (4.2.14), and (4.2.16) respectively into the agent's utility function (4.2.8) and differentiating it with respect to τ (the derivation is outlined in Appendix G). This results in the following optimal solution for τ:

$$\tau^* = \frac{\gamma\theta a}{(1+\beta+\gamma\theta)(1-ab)} \tag{4.4.1}$$

For the solution to τ^* to be interior, the condition $b < \frac{1}{a}$ must be satisfied. When $b > \frac{1}{a}$ a corner solution where $\tau^* = 0$ occurs. This leads to $h_{t+2} = 0$. On the other hand, if $b < \frac{-\theta\gamma}{(1+\beta+\gamma\theta\beta)} + \frac{1}{a}$, $\tau^* = 1$. However, in the range $p \le b < n$, which was established previously, as required to ensure that the solutions to the decision problem are interior and the elasticity of substitution at steady state is greater than 0, neither of these corner solutions will emerge. Thus, from (4.3.1), in the range $p \le b < n$, a higher value of *b* results in a higher value of τ^*. Thus:

Proposition 4: *The tax rate which enables the agent to maximise utility is increasing in the value of b and hence the elasticity of substitution.*

Proposition 4 captures the idea that when public and private education inputs are closely substitutable, people are willing to pay more tax and strengthen the public education system. Given that public education expenditure is the essential input in the education production function, in a setting where both public and private education expenditures provide a similar benefit to the child, parents are willing to pay more tax and improve the public education system, as it is the relatively more important input, given that when public education is 0, the entire education production collapses to 0, while a certain level of education can still be produced even when private education is 0, as long as $b > 0$. If the public education system is of a poor quality, it would be an inefficient substitute for private education, and this would encourage parents to supplement public education with private education services. In such an instance, parents would not be willing to pay high taxes towards the maintenance of the inefficient public education system, as this money would be better spent on buying high quality private educational services for the child.

4.5 CONCLUSION

We develop an overlapping generations model to explore how public and parental expenditures impact long run macroeconomic outcomes. Using a variable elasticity of substitution "education production function," in which we assume public education to be the essential input, we show analytically that steady state utility, human capital, physical capital, output, and the human to physical capital ratio are increasing in the aggregate substitutability parameter b, and hence, the elasticity of substitution. This section presents the concluding remarks relating to these results and provides policy recommendations.

The model's prediction that a higher elasticity of substitution between private and public education expenditures yields better macroeconomic outcomes is of practical relevance. A higher elasticity of substitution implies that the two types of expenditure are perceived as largely substitutable for one another, which means that even though public education has a stand-alone existence, there is a private sector that provides comparable services. As higher aggregate substitutability reduces parental expenditure on education, it also entails important implications in terms of access. Especially in developing countries, if public and private education are of comparable quality, it would encourage more children to attend public schools and thereby reduce parental expenditure on education. In a setting where the two inputs are less substitutable for one another, sending children to public schools is likely to entail complementary private expenditures, such as out-of-school tutoring and facility fees.

Therefore, the results suggest that policymakers should aim to develop private participation in the education industry by offering greater support, possibly in the form of tax benefits, training educators in the private sector, and developing a common curriculum that can be used by both the public and private sectors. The quality of private provision can be ensured through arrangements such as registration, quality audits, and reviews undertaken by government authorities that provide parents with the confidence to consider the private education sector to be a close competitor to the public sector. Ultimately, developing public and private education that are of comparable quality is about giving parents a greater choice with regard to the type of education they wish to provide to their children.

There has been a move towards increased private funding of education in many countries in recent years (Levin & Belfield, 2003). Increasing private sector participation in the education industry and intense competition in the job market has led to a rise in the expenditure incurred by parents for educating their children in many countries. For instance, private expenditure accounted for 12.2% of total educational expenditure in the OECD in 2000, but this percentage rose to 16% by 2009 (OECD, 2012a). Encouraging such increased private participation could encourage the public sector to be more competitive in the provision of education and increase the degree of uniformity between public and private education expenditures, yielding greater choice for parents and improving long run macroeconomic outcomes as a result.

As we do not explicitly include any quality considerations in the model, it is useful to consider our results in conjunction with a large volume of interdisciplinary empirical literature that evaluates the relative effectiveness of private education as opposed to public education. While a large number of empirical studies demonstrate that private schools outperform public ones in terms of student performance and efficiency (see Coleman, Hoffer, & Kilgore, 1982; Jimenez & Lockheed, 1995; Wongsurawat, 2011; Zahid, Muhammad Iqbal, & Tariq, 2011, for instance), there are also some studies that are inconclusive on whether attending private schools leads to an improvement in student achievement (see, for example Alexander & Pallas, 1983; Sander, 1999). Thus, while the present study yields some interesting insights into the public-private mix of education expenditures, further extensions of the model presented in this paper may be able to yield deeper insights into these issues. In particular, as we consider a framework in which intra-generational heterogeneity is absent, we do not consider the impacts on inequality associated with greater aggregate substitutability. Distributional considerations are, however, an important dimension of this issue, and consequently an important direction of future research.

The fact that the ratio of physical to human capital rises with the degree of aggregate substitutability is also an interesting result with implications for inter-generational redistribution, as in our model, middle-aged agents own human capital and the old agents own physical capital. Therefore, despite a higher aggregate substitutability between public and private education expenditures directly contributing towards strengthening the stock of human capital, the model suggests that by helping to raise savings, it provides a greater contribution towards physical capital accumulation in the economy. As suggested by Bertola (1996), modelling these considerations in a continuous time OLG framework may enable one to explore these intergenerational redistribution concerns in greater depth. Thus, exploring both intra and intergenerational aspects of the model within a continuous time OLG framework might be a further extension of the model.

Chapter 5: Health Expenditures and Inequality: a Political Economy Perspective

5.1 INTRODUCTION

Public expenditure on health care is often considered to be a means of achieving equitable and efficient outcomes for an economy due to its redistributive nature. However, individuals can also undertake *private* investments in education and health. As suggested by Chakraborty and Das (2005), Deaton (2003), and Ray and Streufert (1993), unequal private investments in health, which arise from inequalities in the distribution of income and wealth, could lead to a persistence of income and wealth inequality in the long run. Given the importance of both public and private health expenditures, several extant studies have explored the manner in which the mix of these expenditures can impact macroeconomic outcomes (see for instance, Bhattacharya & Qiao 2007; Lahiri & Richardson 2009; Li et al. 2012; Varvarigos & Zakaria, 2013).

In this paper, we examine how the public-private mix of expenditures on health contributes towards mortality and wealth in the long run in the context of an overlapping generations model comprising of heterogeneous agents. Both public and private investments in health contribute towards the 'production' of an agent's health, and consequently her mortality. Public expenditures on health in this economy are endogenous; they are determined by a political economy mechanism, whereby agents vote on the proportion of tax revenue to be used to finance public health care. The remainder of the revenue is allocated equally among agents in the form of a lump sum transfer. A key innovation in our approach, relative to previous studies, is our focus on the distinction between individual and aggregate level factors that impact on an agent's health.

Specifically, central to this paper is a "health production function" of the variable elasticity of substitution (VES) form, discussed in Revankar (1971), in which public and private expenditures on health are inputs that contribute towards an agent's health and longevity.[27] The VES form, in addition to having some intuitively appealing and relevant properties in the context of health, allows the degree of substitutability between public and private expenditures, as measured by the elasticity of substitution between them, to vary between agents.[28] An agent's elasticity of substitution is determined by her wealth and a parameter of the production function, labelled "b", which we interpret as an *aggregate* measure of the *ease of substitutability* between public and private inputs. This means that the elasticity of substitution varies *between agents* and *across time*, introducing a richer set of dynamics that feed into the determination of the political economy outcomes and long run results.

As the notion of substitutability discussed above is germane to the interpretation of our framework and its results, it is worthwhile elaborating upon this point. In particular, we suggest that the "aggregate substitutability" parameter is conditioned by a number of institutional parameters, such as the quality of public health care and the regulatory framework governing the economy. For example, if the quality of public health care is inferior relative to private health care, then it is obvious

[27] The VES production function discussed in Revankar (1971) is of the standard, technological form that combines labour and capital as inputs to produce output.

[28] Our notion of "substitutability", given that it enters the preference side of the economy, refers primarily to a personal trade-off between two types of goods (public and private health inputs), rather than substitutability in the sense of a market-induced trade-off created due to the differences in the price levels of the goods. However, the market trade-off enters in an implicit sense, as an agent's wealth level impacts on her elasticity of substitution between the two goods.

that there is low substitutability between these services.[29] Furthermore, if regulation prohibits the private provision of certain health goods and services, there is lower substitutability at the aggregate level in the sense that goods and services available under the public health care system have no private counterpart.[30] In general, however, the degree of aggregate substitutability is likely to be relatively higher in developed countries, where public and private health services are of a comparable quality.

In relation to the individual level, Gouveia (1997), for example, suggests that an individual's ability to substitute between private and public health inputs is a perception conditioned by several micro-level factors, which are likely to differ between various individuals, depending on their specific circumstances. These micro-level factors could include income/wealth, location, and socio-economic status.[31] In general, however, if the "aggregate substitutability" between these inputs is high, the individual substitutability will also be high. For example, if public and private health services were of a comparable quality and affordability, and had good coverage across locations, the aggregate substitutability would be high, and this would have a positive impact on individual substitutability. However, if they were not of comparable quality, with substantial differences in the prices and coverage across locations, aggregate substitutability would be low, and it would worsen individual substitutability.

[29] For instance, Filmer, Hammer and Pritchett (2000) refer to studies from Sri Lanka, Pakistan, Indonesia, and El Salvador that reveal a tendency for patients to "bypass" public facilities with poor equipment, inadequate staff, and a low level of accountability and efficiency in favour of private health care facilities capable of providing a higher quality of service, which reduces the aggregate substitutability between the two types of expenditures.

[30] In Canada, for example, the private sector is prohibited from providing surgery at a profit (Lülfesmann & Myers, 2010). This artificially restricts the degree of substitutability between public and private health expenditures.

[31] For instance, in many developing countries, the better equipped state hospitals and schools are likely to be located in the cities. In such a context, people living in rural areas may be compelled to seek private medical services due to the inability to travel long distances to the cities. Thus, a person's place of residence could restrict his/her ability to substitute between public and private health expenditures.

Key results of this paper—analytical and numerical—suggest that a higher degree of aggregate substitutability yields better long run outcomes in the form of higher average wealth levels and lower inequality. This occurs *through* the political economy mechanism—in the presence of higher aggregate substitutability, agents vote for a higher proportion of revenue to be allocated towards improving public health care. Given that aggregate substitutability between public and private health inputs is likely to be higher in developed countries, this result is of empirical relevance, as public health spending is much higher in developed countries when compared to developing ones. Taking into consideration the merits associated with a higher degree of aggregate substitutability, an obvious policy implication is that institutional reforms directed at improving aggregate substitutability will lead to better macroeconomic outcomes. Another interesting result of the paper is that a higher value of another parameter a, which can be interpreted as the 'pure' share of public health in the health production function, is also associated with higher average wealth levels and lower inequality in the long run. The underlying rationale for this result is that, in a setting where a person's health capital is determined by both public and private health expenditures, greater efficiency of public health care results in individuals having to undertake lower private health expenditure, thereby enabling them to leave more bequests for their offspring.

The importance of these results is further motivated by the intuitive appeal of the VES construct, which, as discussed above, enables the degree of substitutability to be governed by aggregate factors, as well as the economic circumstances of the individual. Secondly, the outcome of this essay, that a greater "ease of substitutability", as measured by a higher elasticity of substitution between public and private health expenditures, leads to better aggregate outcomes, is also intuitively plausible. The parameter b, which captures all the aggregate-level factors that affect the ease with which an individual can substitute between private and public health care expenditures, is an indicator of the range of choices available to *all* agents in the economy at the aggregate level. This range is likely to be larger if goods and services provided by the public and private sectors are of comparable quality and availability. While the aggregate quality, efficiency, and accessibility of both the private and public components of any health care system are certainly important, ignoring individual micro-level heterogeneity amounts to ignoring elements that are of relevance in determining the overall effectiveness of the health care system, which in turn has a bearing on health inequality and macroeconomic performance. The VES form, due to its ability to distinguish between the individual and the aggregate, is therefore a more realistic framework to address issues in the health macroeconomics literature.

However, as recognised in the political economy literature, policies and institutions are essentially endogenous; economic agents influence their determination, either through the presence of vested interest groups, or by means of a voting mechanism (see for example, Alesina & Rodrik, 1994; Krusell & Rios-Rull, 1996). To address this issue, we compare the proportion of tax revenue allocated towards public health care, determined through the political economy process, with that chosen by a benevolent social planner whose objective is to maximise the collective welfare of all agents in the economy. A natural question that arises is why we select the proportion of tax revenue allocated towards public health care as the parameter which is determined through a political economy mechanism rather than considering the parameter b as the one agents vote on. While this matter is examined in greater depth in Section 5.4, it should be noted here that the parameter b is a deeply ingrained institutional parameter that is difficult to measure. Furthermore, developing a functional form for b that enables it to be determined through the political economy process is difficult, as it is not possible to infer the manner in which the value of b is affected by government spending. Hence, considering the proportion of tax revenue allocated towards public health care as the parameter which is determined through a voting process, rather than the parameter b, is a more realistic approach from a political economy perspective.

We see that for a substantial range of parameter values, the political economy outcome converges to the welfare maximising outcomes under both the utilitarian and Rawlsian approaches in the long run, albeit it can be sub-optimal on the transition path. Even in the instances where the political economy outcome and the welfare maximising outcomes do not coincide, the winning proportion of tax revenue allocated towards public health care is always equal to either the utilitarian or Rawlsian social welfare maximising value, suggesting that the determination of the proportion of tax revenue allocated towards health care through a voting process is largely socially desirable in this context. During transition, we see that the winning proportion of tax revenue allocated towards health care is closer to the welfare maximising proportion under the utilitarian approach than that under the Rawlsian approach.

The utilitarian outcome, which weighs all agents' utilities equally, is likely to be closer to the political economy outcome, which weighs each agent's *vote* equally. The Rawlsian outcome attaches the highest weight to the utility of the poorest agent. The poor agents in this economy prefer redistribution in the form of the lump sum transfer; given that it is a perfect substitute for consumption, it is a necessity for all agents in the economy. Hence, poor agents are likely to prefer a bigger share to tax revenue to be allocated towards the lump-sum transfer, which enables them to enjoy higher consumption. Considering the utilitarian, Rawlsian, and political economy paradigms in the context of the issues addressed above is another useful contribution of the paper, as it broadens the scope of the framework to provide potential explanations for the diversity in health outcomes across different countries. In practice, the determination of public health expenditure could occur through political economy or social planner focussed approaches. While in reality, social planning approaches that exactly resemble either the Rawlsian or the utilitarian viewpoints do not exist, these paradigms provide the outer limits of a spectrum within which such approaches may fall.

This essay is organised as follows: Section 5.2 provides a background and motivation for the study where we present the VES health production function and discuss its features. Section 5.3 introduces the benchmark model and presents analytical results. Section 5.4 looks at a political economy extension to the benchmark model. Section 5.5 presents the results of the numerical experiments and the related discussion, and finally Section 5.6 comprises of concluding remarks and policy implications. A number of proofs and derivations are provided in the Appendices.

5.2 BACKGROUND AND MOTIVATION

Contemporary developments in medicine have bestowed upon people an enhanced ability to influence their health, and as a consequence, individuals are increasingly regarded to be the ultimate producers of their own health, which is the essence of the concept of the "health production function" (Zweifel, Breyer, & Kifmann, 2009, pp 75-76). The health production function is a concept introduced to mainstream economics by Grossman (1972) to capture the preferences of individuals towards factors that have an impact on their health (Johannesson, Johansson, Kriström, Borgquist, & Jönsson, 1993). Guided by this idea, we use a simplified health production function of the VES form in which public and private health expenditures are the inputs needed for creating a person's health capital. Attempts in the extant literature to measure the idea of substitutability between public and private health inputs in the manner we interpret it are limited. The use of the VES health production function enables us to capture the idea of aggregate substitutability, as well as *individual* variations in the elasticity of substitution, which is an aspect extant studies such as Bhattacharya and Qiao (2007) and Lahiri and Richardson (2009) have not taken into consideration. In order to explain this point, we first provide a brief overview of these studies.

Bhattacharya and Qiao (2007) develop an overlapping generations model with homogeneous agents within a particular generation, where an agent's survival probability in old age is a function in which public and private investments aimed at enhancing longevity appear as interactive, complementary terms. The authors justify this formulation by suggesting that the marginal increase in longevity resulting from additional out-of-pocket expenditure on health rises when the state health care system is bigger and more efficient. However, the types of complementary private inputs discussed by these authors usually entail out-of-pocket expenditures, which may not be affordable to a majority of the population in developing countries. Hence, whilst complementarity between public and private inputs might be a reasonable assumption at the macro-level in the context of a developed country, it is not necessarily appropriate as an *a priori* assumption for developing or transitional economies.

On the other hand, Lahiri and Richardson (2009) develop an overlapping generations model with heterogeneous agents, with a constant elasticity of substitution (CES) health production function. The authors explore the long run wealth distributions and the political economy outcomes associated with different values of the elasticity of substitution. The health production function used by these authors offers greater flexibility when compared to the longevity production function used in

Bhattacharya and Qiao (2007), as it allows the exploration of the macroeconomic impacts associated with a range of elasticity of substitution values. Li et al. (2012) also develop a similar model with a CES health production function to explore the public-private mix of health care expenditures.[32] However, it is difficult to distinguish between aggregate and individual substitutability using a CES health production function, as the elasticity of substitution parameter in such a function is constant across individuals. As such, the meaning of substitutability in our paper is distinct from the notion of substitutability in the above-mentioned papers.[33]

In light of the above discussion, we can now examine how to distinguish between aggregate and individual substitutability using a VES health production function, which takes the following form:

$$h_t = \left(h_t^G\right)^a \left(h_t^P + abh_t^G\right)^{1-a} \qquad (5.2.1)$$

where $0 \le a \le 1,$ and $\left(\dfrac{h_t^G}{h_t^P}\right)^{-1} \ge -b$

In (5.2.1), h_t is an individual's health capital, h_t^G is public health spending per person and h_t^P is private health expenditure incurred by an individual. The parameter restrictions given above are required to ensure that (5.2.1) satisfies the standard properties of a production function. In the VES form, public expenditure is the essential input, because if $h_t^G = 0, h_t = 0.$ Such an assumption may generally be

[32] To our knowledge, the first attempt at examining the long-run impact of substitutability between public and private health inputs using a CES construct for the health production is by Lahiri and Richardson (2009). Li et al (2012) follow a very similar approach, albeit with a different set of objectives.

[33] We emphasise that this difference is not only of relevance to the meaning and interpretations of the features of our model, but also its results. For example, Lahiri and Richardson (2009) show that greater substitutability, as per their interpretation, can have harmful implications for inequality.

justifiable on the grounds that public spending on health care is instrumental towards the improvement of a person's life expectancy (Glomm & Ravikumar, 1997).[34]

The elasticity of substitution between public and private inputs is:

$$\eta = 1 + b\frac{h_t^G}{h_t^P} \tag{5.2.2}$$

In (5.2.2), the parameter b captures the degree of aggregate substitutability, which has a direct influence on the individual's elasticity of substitution. Although the values of b and h_t^G are common to all individuals, the value of h_t^P would differ between agents as it is positively related to wealth in our model. If $b > 0$, wealthier individuals have a lower elasticity of substitution than their poorer counterparts, but if $b < 0$, the elasticity of substitution of wealthier individuals would be higher.

As emphasised earlier, the meaning of substituability between public and private health inputs varies depending on the framework in question. Subject to this caveat, extant literature provides different perspectives on this issue. For instance, Bhattacharya and Qiao (2007) suggest that in developed countries where people are generally wealthier, public and private health expenditures are likely to be characterised by a greater degree of complementarity. On the other hand, Bidani and Ravallion (1997) argue that wealthier people find it easier to substitute public health care with privately provided alternatives and would therefore have a higher elasticity of substitution. Fabbri and Monfardini (2009) also support the latter assertion by

[34] In the case of health, considering public expenditure as the essential input is reasonable, as issues such as ageing populations and obesity in developed countries, and disease, war, and malnutrition in the case of many developing countries are challenges the private sector cannot typically address. Furthermore making public expenditure the essential input ensures that even the poorest members of society, who are unable to incur much private health expenditure, would be able to enjoy a fair level of health capital.

noting that richer people are in a better position to substitute public health care with private health services due to their ability to afford private health insurance. Hence, by varying the value of b in the health production function we can capture both of these possibilities suggested in the literature.

Another interesting feature associated with equation (5.2.2) is that an individual's elasiticity of subsitution between public and private health expenditures is affected by public health spending per person, h_t^G. Generally, public health spending per person is positively related to an economy's level of development. This is illustrated in Figure 5.1, which shows the public health spending as a percentage of total health spending is higher in regions comprising of developed countries. This is also demonstrated in Table 5.1, which shows that although public spending on health per capita has increased over time in all the country groups considered, it is higher in regions where countries are more developed.

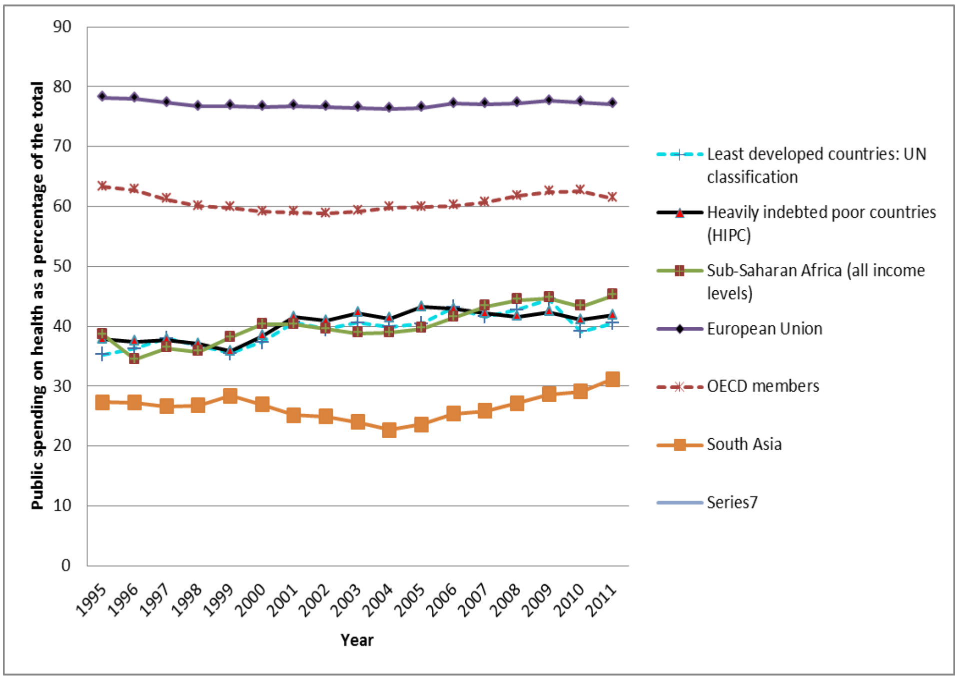


Figure 5.1: Public spending on health as a percentage of total health expenditure

Source: The World Bank

Table 5.1: Public spending on health care per capita

Region	1995	1999	2003	2007	2011
Sub-Saharan Africa	38.03	33.48	42.03	72.18	94.60
Heavily indebted poor countries	14.01	16.08	17.59	33.25	47.19
Least developed countries	11.37	11.58	14.11	26.86	42.05
OECD members	2097.44	2237.04	2913.05	3866.02	4593.01
European Union	1646.05	1620.59	2165.89	3188.71	3530.23

Source: World Bank data

The above stylised features of public health spending provide some further motivation for considering the VES construct of this paper. Firstly, in the context of health, the data is suggestive of a strong link between individual and aggregate health outcomes. The VES provides a way of explaining this link and interpreting the mechanisms underlying it.

Furthermore, we wish to emphasise that unlike in the case of a traditional production function, where the elasticity of substitution between labour and capital at the firm level is primarily determined by the technology used in the production process, the determinants of an individual's elasticity of substitution between public and private expenditures in the health production function have a bearing upon the *preferences* of an individual. In short, one has to appreciate that the health production function differs from a traditional production function as "health production" in the framework to follow does not take place in a market—it is something that is "produced" as an end result of an agent's and the government's expenditures on health goods and services/health care systems. Bearing this in mind, we look the benchmark model in the next section.

5.3 THE BENCHMARK MODEL

We consider a two period overlapping generations model where each generation consists of N agents. Time is discrete and given by $t=1,2,...$. All agents safely survive through youth, but an agent's probability of survival into old age is determined by health investment in youth, which in turn comprises of private health expenditure undertaken by the agent during youth, as well as health care services provided by the state. During youth, the agent gives birth to a single offspring, whose wealth endowment, as in Chakraborty and Das (2005), depends on whether her parent survives safely through old age or not. In the former case, the parent leaves bequests for her child, while in the latter case, the child inherits her parent's savings, which can be considered as unintended bequests. We assume an imperfect wealth market such that interest on savings can be enjoyed by the individual only if she survives safely through old age. If the parent dies prematurely, her child inherits only the gross value of her savings as an unintended bequest.[35] Public health care services are financed by a proportional tax of rate τ imposed on the heterogeneous wealth endowments W_t of the young agents in the economy. These wealth endowments have a probability density function $g_t(W_t)$ in the support $[0,\infty]$, so that average tax revenue collected by the

[35] Such unintended or accidental bequests arise only in the absence of perfect annuities markets. In an environment where lifetime is uncertain, agents save in order to finance consumption in the event they live to reach a very old age (Abel, 1985; Strawczynski, 1993). Yaari (1965), whose work pioneered a large volume of research on bequests, uses the term "actuarial note" to capture the concept of a life insured annuity that pays the buyer an allowance during her lifetime and is cancelled off once she dies, and shows that, in the absence of altruism, an agent would annuitise all her savings. In this model, we assume that if the parent dies prematurely, the child inherits the gross value of her parent's savings without interest, which requires the assumption that annuities markets, although present, are imperfect, as interest accumulated on savings can only be enjoyed by the individual and not by her offspring.

government in period t is $\tau\int_0^\infty W_t g_t(W_t)dW_t = \tau\overline{W_t}$. Of this average tax revenue, the government allocates a proportion ψ towards the provision of equitable public health care and the remainder is paid to the young individuals as a lump-sum transfer.

To improve analytical tractability, and without any loss of generality, we ignore youthful consumption, so that an individual derives utility from consumption and bequests left to her offspring. Thus, in period t, the agent's utility function is:

$$U_t = \phi(h_t)[U(c_{t+1}) + \theta V(m_{t+1})] \tag{5.3.1}$$

where c_{t+1} is old age consumption, m_{t+1} denotes bequests, $\theta > 0$ is a parameter capturing the degree of parental altruism, and $\phi(h_t)$ represents the probability of surviving to old age, which depends on the individual's health capital h_t. As mentioned in Section 5.2, this "health" output is determined by both public and private investments in health and is captured by the VES health production function, which is reproduced below for convenience:

$$h_t = \left(h_t^G\right)^a \left(h_t^P + abh_t^G\right)^{1-a} \tag{5.3.2}$$

As in Chakraborty and Das (2005), we assume that the survival probability $\phi(h_t)$ has the following functional form:

$$\phi(h_t) = \begin{cases} \chi h_t^{\varepsilon} \text{ for } h \in \left[0, \hat{h}\right] \\ \bar{\phi} \text{ otherwise,} \end{cases}$$

where $\bar{\phi} \leq 1$, and χ and ε are constants with $\chi \geq 0$ and $0 \leq \varepsilon \leq 1$

(5.3.3)

The above survival probability function is concave in health capital h_t for values below $\bar{\phi}$, which is the maximum survival probability an agent is capable of achieving, but additional health capital cannot increase one's survival probability once $\bar{\phi}$ is reached. As such, unlike a strictly concave survival probability function of the type presented in Chakraborty (2004) and Leung and Wang (2010), a survival probability function of this form allows for heterogeneity in health capital to arise from the initial distribution of wealth, because in a particular time period, there are some people with health capital below $\hat{h}$ who achieve a survival probability in the concave section of the survival probability function, while there are others who achieve the maximum survival probability of $\bar{\phi}$ by accumulating a health capital equal of $\hat{h}$ where $\hat{h} = \left(\frac{\bar{\phi}}{\chi}\right)^{\frac{1}{\varepsilon}}$. Furthermore, we later show that this survival probability function, when combined with the VES health production function, results in a bimodal wealth distribution.

As everybody has access to the same value of public health care services, an agent has to adjust her private health expenditure h_t^P so as to attain the health capital $\hat{h}$ associated with the maximum survival probability $\bar{\phi}$. Denoting this critical level of private health expenditure by $\hat{h}_t^P$ we have:

$$\hat{h}_t^p = \left[\frac{\left(\frac{\bar{\phi}}{x}\right)^{\frac{1}{\varepsilon}}}{\left(\psi\tau\overline{W}_t\right)^a}\right]^{\frac{1}{1-a}} - ab\psi\tau\overline{W}_t \tag{5.3.4}$$

Recall that the parameter b in the above expression comes from the VES health production function (5.3.2). In our model, we refer to this parameter as representing "aggregate substitutability". It can be seen that a higher value of the aggregate substitutability parameter b causes the critical level of private health spending to decline. In other words, if public and private health expenditures are closely substitutable, it reduces the incentive for an individual to undertake private health spending.

Given heterogeneous wealth endowments, agents with wealth endowments below a critical level $\hat{W}_t$, who cannot afford the critical level of private health expenditure $\hat{h}_t^P$, are located on the concave section of the survival probability function presented in (5.3.3). On the other hand, those with wealth levels greater than or equal to $\hat{W}_t$ who can afford the critical private health expenditure of $\hat{h}_t^P$ are capable of achieving the maximum survival probability of $\bar{\phi}$. This critical wealth level $\hat{W}_t$ is derived in Appendix K. For a complete characterisation of optimal choices, we need to consider the optimal decisions of agents with wealth levels below $\hat{W}_t$ as well as those with wealth levels above $\hat{W}_t$.

The agent maximises (5.3.1) subject to the following budget constraints:

$$h_t^p + s_t = \overline{w_1} + W_t(1-\tau) + \tau(1-\psi)\overline{W_t} \tag{5.3.5}$$

$$c_{t+1} + m_{t+1} = \overline{w_2} + \bar{R}s_t \tag{5.3.6}$$

Equation (5.3.5) above is the budget constraint the agent faces in youth. The individual inelastically supplies one unit of labour to earn a subsistence wage of $\bar{w}_1$ in youth and inherits a wealth of W_t from her parent, of which, as described above, a proportion τ is charged by the government as tax. Out of the government's tax revenue a proportion ψ is allocated towards the provision of public health services, which results in each agent receiving public health care with an approximate value of $\psi\tau\overline{\mathrm{W}}_t$, while the remaining tax revenue is divided equally amongst the population in the form of a transfer. During youth, the individual divides her total after-tax earnings between private health inputs h_t^P and savings s_t. In old age, the agent uses her subsistence wage of $\overline{w}_2$ and her savings, which have accumulated a gross interest of $\overline{R}$ to finance consumption and bequests. As explained before, if the parent survives safely through old age, the child's wealth is equal to the intended bequest from the parent, so that $W_t = m_t$, and if the parent dies prematurely the child inherits the gross value of her parent's youthful savings, so that $W_t = s_{t-1}$. The agent's utility maximisation problem involves maximising (5.3.1) subject to equations (5.3.2) to (5.3.6). We assume that the agent's utility functions are of the familiar iso-elastic form. Thus, $U(c_{t+1}) = \frac{c_{t+1}^{1-\sigma}}{1-\sigma}$ and $V(b_{t+1}) = \frac{b_{t+1}^{1-\sigma}}{1-\sigma}$ where $0 < \sigma < 1$.[36] We assume a small open economy setting where the interest rate is exogeneously given.

For agents with wealth levels $W_t \le \hat{W}_t$, the first order conditions are:

$$U'(c_{t+1}) = \theta V'(d_{t+1}) \tag{5.3.7}$$

$$\frac{\partial\phi}{\partial h_t} \times \frac{\partial h_t}{\partial h_t^P}\left[U(c_{t+1}) + \theta V(d_{t+1})\right] = \overline{R}\phi(h_t)U'(c_{t+1}) \tag{5.3.8}$$

[36] This type of cardinal assumption is typically made in models with endogenous time preference. While Chakraborty and Das assume $0 < \sigma < 1$, others, such as Lahiri (2002), assume it to be negative.

Equation (5.3.7) captures the fact that the marginal utility from consumption needs to be equal to the marginal utility from bequests at utility maximisation. Equation (5.3.8) captures the need for the marginal benefit of spending on private health services measured in terms of a higher survival probability to be equal to the marginal cost, which is the potential old-age consumption she foregoes.

As an agent with a wealth level $W_t > \hat{W}_t$ incurs private health expenditure of $\hat{h}_t^P$ and achieves the maximum survival probability of $\bar{\phi}$, the only first order condition associated with her utility maximisation problem is given by (5.3.7).

Upon setting $\beta = \theta^{-1/\sigma}$, $\gamma = \dfrac{1-\sigma}{\varepsilon(1-a)}$ and $Y_t = \bar{w}_1 + (1-\tau)W_t + \tau(1-\psi)\overline{W}_t + \dfrac{\bar{w}_2}{\bar{R}}$, the analysis of both sets of first order conditions yield the following closed form solutions for consumption, savings, bequests, and private health expenditure for the entire wealth distribution:

$$c_{t+1}^{*} = \begin{cases} \dfrac{\beta\gamma\bar{R}\left(Y_t + ab\psi\tau\overline{W}_t\right)}{(\beta+1)(1+\gamma)} & \text{for } W_t \le \hat{W}_t \\[2ex] \dfrac{\beta\bar{R}\left(Y_t - \hat{h}_t^P\right)}{(\beta+1)} & \text{for } W_t > \hat{W}_t \end{cases} \tag{5.3.9}$$

$$m_{t+1}^{*} = \begin{cases} \dfrac{\gamma\bar{R}\left(Y_t + ab\psi\tau\overline{W}_t\right)}{(\beta+1)(1+\gamma)} & \text{for } W_t \le \hat{W}_t \\[2ex] \dfrac{\bar{R}\left(Y_t - \hat{h}_t^P\right)}{(\beta+1)} & \text{for } W_t > \hat{W}_t \end{cases} \tag{5.3.10}$$

$$s_t^* = \begin{cases} \dfrac{\gamma\left(Y_t + ab\psi\tau\overline{W}_t\right)}{(1+\gamma)} - \dfrac{\overline{w}_2}{\overline{R}} & \text{for } W_t \leq \hat{W}_t \\ Y_t - \dfrac{\overline{w}_2}{\overline{R}} - \hat{h}_t^P & \text{for } W_t > \hat{W}_t \end{cases} \tag{5.3.11}$$

$$h_t^{P*} = \begin{cases} \dfrac{Y_t - \gamma ab\psi\tau\overline{W}_t}{(1+\gamma)} & \text{for } W_t \leq \hat{W}_t \\ \hat{h}_t^P & \text{for } W_t > \hat{W}_t \end{cases} \tag{5.3.12}$$

$$\text{where } \hat{W}_t = \frac{\left[\dfrac{\left(\dfrac{\overline{\phi}}{x}\right)^{\frac{1}{\varepsilon}}}{A\left(\psi\tau\overline{W}_t\right)^a}\right]^{\frac{1}{1-a}} (1+\gamma) - \tau\overline{W}_t\left[(ab-1)\psi + 1\right] - w_t - \tau(1-\psi)\overline{W}_t - \dfrac{\overline{w}_2}{\overline{R}}}{(1-\tau)} \tag{5.3.13}$$

From (5.3.9) to (5.3.12) we have:

Proposition 1: *For all agents, consumption, bequests and savings are increasing in b, while private health spending is decreasing in b.*[37]

Proposition 1 above captures a key characteristic of the model and is the basis for interpreting most of the results that follow. The intuition behind Proposition 1 is that a higher value of the aggregate substitutability parameter b implies a greater degree of *similarity*, in a qualitative sense, between public and private health expenditures. Hence, this would result in an agent needing to undertake lower private

[37] It is obvious that this is the case, as b enters linearly into all the expressions above. Note also that $\hat{h}_t^P$ is given by (3.4) above, which establishes that private health spending both above and below the critical level of wealth is decreasing in b.

expenditure on health to achieve a certain level of health capital and thereby survival probability. Lower private health expenditure allows the agent to save more and enjoy more consumption and leave more bequests in old age. Thus, an agent's utility is clearly increasing in b.

Somewhat heuristically, this observation is analogous to the intuition in the context of a general production function in which a higher elasticity of substitution between the two inputs leads to a lower cost of production. Figure 5.2, in which η represents the elasticity of substitution, presents a standard, "textbook" comparison of the linear, Cobb-Douglas and Leontief production functions. It is clear that the linear production function, which has an infinite elasticity of substitution, is associated with the lowest cost of production. This is because a higher elasticity of substitution enables the firm to produce the assumed output with the least inputs as a higher degree of substitutability between inputs allows the firm to save costs by using the cheaper input. Relating this observation to the context of the health production function, the lower expenditure on private health services resulting from a higher value of the aggregate substitutability parameter b represents a similar "cost saving" for the agent, which she can utilise on consumption and bequests.

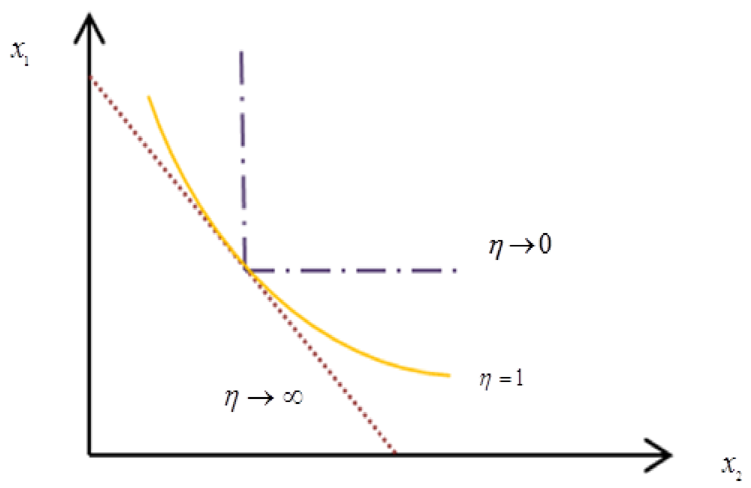


Figure 5.2: Comparison of the effects of the elasticity of substitution

Going back to our analysis of equations (5.3.9) to (5.3.12), the effect of average wealth $\overline{W_t}$ on the optimal values of the decision variables is less clear. For agents with wealth levels $W_t \leq \hat{W}_t$, if $b > \frac{-(1-\psi)}{a\psi}$, c_{t+1}^*, m_{t+1}^* and s_t^* increase when $\overline{W_t}$ increases. Furthermore, for these agents, $h_t^{P^*}$ and $\overline{W_t}$ are positively related if $b < \frac{(1-\psi)}{a\psi\gamma}$, and negatively related otherwise. For agents with wealth levels $W_t > \hat{W}_t$, it is clear when $b > 0$, consumption, bequests and savings are increasing in $\overline{W_t}$ while private health spending is decreasing in $\overline{W_t}$ since $\frac{\partial \hat{h}_t^P}{\partial W_t} < 0$ given $b > 0$. Hence, the utility of *all* agents is increasing in average wealth $\overline{W_t}$ as long as $b > 0$, implying that the ability for average wealth levels to contribute towards the improvement of individual utility is crucially dependent upon a high aggregate substitutability parameter.

Recall that an individual inherits a bequest if her parent survives safely through old age, but if her parent dies prematurely, she inherits an unintended bequest, which is her parent's savings. Hence, we can provide the following characterisation for intended and unintended bequests for the entire wealth distribution:

$$m_{t+1}^* = \Omega_{t+1}^1(W_t) \equiv \begin{cases} \dfrac{\gamma\overline{R}\left(Y_t + ab\psi\tau\overline{W_t}\right)}{(\beta+1)(1+\gamma)} & \text{for } W_t \leq \hat{W}_t \\ \dfrac{\overline{R}\left(Y_t - \hat{h}_t^P\right)}{(\beta+1)} & \text{for } W_t > \hat{W}_t \end{cases} \tag{5.3.14}$$

$$s_t^* = \Omega_{t+1}^2(W_t) \equiv \begin{cases} \dfrac{\gamma\left(Y_t + ab\psi\tau\overline{W_t}\right)}{(1+\gamma)} - \dfrac{\overline{w_2}}{\overline{R}} & \text{for } W_t \le \hat{W}_t \\ Y_t - \dfrac{\overline{w_2}}{\overline{R}} - \hat{h}_t^P & \text{for } W_t > \hat{W}_t \end{cases}$$

(5.3.15)

Given that for any agent *i*, the parent's probability of surviving into old age is determined by her health capital, the i^{th} agent's expected wealth is given by:

$$\psi_{t+1}^E = W_{t+1}^i = \phi(h_t^i)\Omega_{t+1}^1(W_t^i) + [1-\phi(h_t^i)]\Omega_{t+1}^2(W_t^i) \qquad (5.3.16)$$

It is also important to note that individuals may benefit more from their parents surviving into old age and leaving them intended bequests, as opposed to parents facing an early demise and leaving them unintended bequests (Chakraborty & Das, 2005). This requires that bequests should exceed savings even when income in the second period is 0. In this model, the condition under which this holds true for all individuals in this economy is $\overline{R} > 1+\beta$, which simplifies to the condition in Chakraborty and Das (2005), which is: $\theta > \left(\dfrac{1}{\overline{r}}\right)^{\sigma}$, where $\overline{r}$ is the interest rate.

Chakraborty and Das (2005) also discuss the conditions under which there is a likelihood of polarisation of long run wealth of agents. In our model, we can characterise this result more explicitly, given its analytical tractability.[38] We are, in fact, able to prove the following proposition:

[38] Note that unlike in our paper, in the Chakraborty and Das (2005) framework, bimodality is not inevitable as suggested in their discussion on pages 173-174. In their model, it is an artifact of the shape of the bequest function and whether the bequest function intersects the 45° line in two places rather than 1. This point is discussed further in Appendix L.

Proposition 2: *Long run wealth distributions are bimodal.*

In Appendix L, we prove the existence of two unique steady states for both intended and unintended bequests. The existence of two steady states for both the intended and unintended bequests functions implies that these functions will intersect the 45° line in two places, resulting in a bimodal wealth distribution, as illustrated in Figure 5.3 below:

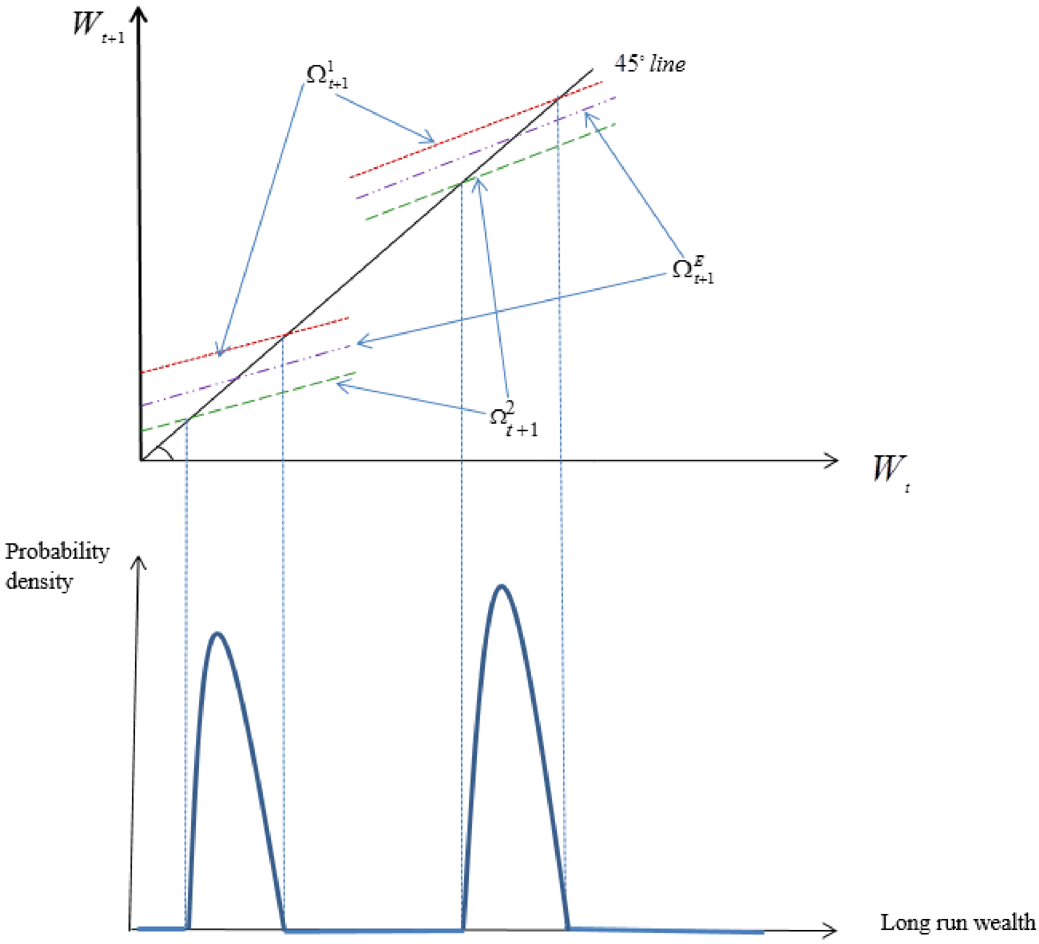


Figure 5.3: Bimodality of long run wealth distributions

The bimodality feature results in the division of individuals into two groups in the long run depending on the initial wealth endowments of their ancestors, and is similar in spirit to the bimodal long run wealth distributions observed in Galor and

Zeira (1993). However, unlike in Galor and Zeira (1993), where twin-peakedness emerges due to heterogeneous investments in education that divides the workforce into skilled and unskilled workers in the long run, in the present model model it occurs through a different mechanism—differences in health investments undertaken by people leads to disparities in bequests left to progeny, yielding a bimodal wealth distribution in the long run.

5.4 POLITICAL ECONOMY EXTENSION

In this section, we consider a political economy extension to the benchmark model introduced in Section 5.3. Decision-making by agents in this economy is now considered to be characterised as a two stage game. At the beginning of each period, there exists a first stage where young agents vote on the value of ψ, which is the proportion of tax revenue to be allocated towards providing public health care (recall from Section 5.3 that the remaining proportion $(1-\psi)$ is divided equally among the agents in the form of a lump-sum transfer). Given that the aggregate substitutability parameter b is the focal point of our discussion, one might expect political economy considerations to be centered around the determination of the the optimal value of b. However, it is important to apppreciate that b is a deeply ingrained "sticky" institutional parameter, which cannot be altered in the short run. Hence, it has a more abstract, non quantifiable nature. Furthermore, most academic and political discussions on institutional reforms in health care are centered around the quality, availability, and accessibility of public health care. One would expect that the population of any country would be able to influence an important parameter such as the proportion of tax revenue allocated towards public health care through the political economy process. As such, our choice of ψ, rather than b, as the parameter determined through a voting process, is more relevant and realistic from a politcal economy perspective. Hence, whilst we stress the need to develop institutional reforms to influence the value of this parameter over the long term, we suggest that the political economy determination of the proportion of tax revenue allocated towards public health care is likely to be an effective means of attaining improvements in agents' utility in the short run.

We assume that only the young agents vote on the value of ψ, because our focus is on intra-generational, rather than inter-generational heterogeneity. Essentially, old agents are not concerned with this vote as they got the opportunity to vote on the proportion of tax revenue they preferred to be allocated towards health care and the lump-sum transfer respectively during youth.[39] Once the winning value of ψ is determined, agents make their consumption and bequest plans in the second stage taking the value of ψ as given. Thus, in Section 3, we considered the second stage of this game.

We now consider the first stage by optimising the indirect utility function of individuals with wealth levels $W_t \leq \hat{W}_t$ and those with wealth levels $W_t > \hat{W}_t$. We obtain this indirect utility function by substituting the optimal values given by (5.3.9) to (5.3.12) in the previous section into the agent's utility function. For both groups of agents, i.e., those below and above the critical level of wealth $\hat{W}_t$, optimising the indirect utility function yields analytical solutions for ψ^*. Proposition 3 characterises the preferred proportion of tax revenue to be allocated towards health care for agents below and above this wealth level.

[39] This assumption is based on the grounds that a given agent votes for public expenditures on health care at the time the health system serves her, which is only during youth. We admit that such a strong assumption restricts us from presenting a complete characterisation of intergenerational heterogeneity, and further work on refining this modelling aspect is likely to be an interesting direction for future research.

Proposition 3: *Let* $q = \frac{1}{a} - \frac{\varepsilon\left[\overline{w}_1 + (1-\tau)W_t + \tau\overline{W}_t + \frac{\overline{w}_2}{R_2}\right]}{\tau\overline{\overline{W}}_t(1+\varepsilon-\sigma)}$. *Then, for individuals with wealth levels* $W_t \leq \hat{W}_t$ *, where* $\hat{W}_t$ *is given by (5.3.13), the preferred proportion of tax revenue allocated towards health care is given by:*

$$\psi^* = \begin{cases} \dfrac{\varepsilon a\left[\overline{w}_1 + (1-\tau)W_t + \tau\overline{W}_t + \dfrac{\overline{w}_2}{\overline{R}}\right]}{\tau\overline{\overline{W}}_t(1-ab)(1+\varepsilon-\sigma)} & \text{if } b < q \\ 1 & \text{if } q < b < \dfrac{1}{a} \\ 0 & \text{if } b > \dfrac{1}{a} \end{cases} \quad (5.4.1)$$

For all agents with wealth levels, $W_t > \hat{W}_t$*, the preferred proportion of tax revenue to be allocated towards health care is given by*:

$$\psi^* = \left(\frac{\overline{\phi}}{x}\right)^{\frac{1}{\varepsilon}}\left[\frac{a}{(1-a)(1-ab)\left(\tau\overline{\overline{W}}_t\right)^{1+a}}\right]^{1-a} \quad (5,4.2)$$

An interesting observation relating to Proposition 3 is that in the range of interior solutions, the preferred proportion of tax revenue allocated towards public health care is increasing in $\frac{W_t}{\overline{\overline{W}}_t}$,which can be regarded as a measure of inequality in society. Therefore, when the extent of inequality in the economy is high, the poorer individuals prefer redistribution through greater public spending on health care. Note that although inequality and the optimal value of ψ for the individual are positively related, the political economy outcome need not be characterised by this feature, as the value of q differs from individual to individual, and the political equilibrium crucially depends on the manner in which q interacts with the parameters in the model. To gain an idea about how these parameters determine the political economy outcome, we need to resort to numerical experiments, which are presented in the subsequent section.

From propositions 3 we have:

Proposition 4: *In the range of interior solutions, for all agents in this economy, the preferred proportion of tax revenue allocated towards health care is increasing in the aggregate substitutability parameter* b.

To understand the rationale underlying Proposition 4, we need to recall that public expenditure is the essential input in the health production function considered in this study. As a higher value of the aggregate substitutability parameter b implies that the two inputs are of comparable quality, people opt to strengthen the component of expenditure they consider to be relatively more important. As mentioned earlier, aggregate substitutability between public and private health inputs is likely to be higher in developed countries. Furthermore, as we noted previously noted, these countries are characterised by relatively higher public health expenditures. Proposition 4 builds a link between these two observations by suggesting that higher aggregate substitutability encourages agents to vote for a higher proportion of tax revenue to be allocated towards public health care.

5.5 NUMERICAL EXPERIMENTS AND ANALYSIS

This section is based on some numerical experiments carried out by varying the values of the aggregate substitutability parameter b and the “pure” share of public spending a. Although we were able to analytically establish that a higher value of the aggregate substitutability parameter b impacts wealth and utility positively in the benchmark model, and individuals' preferred proportion of revenue allocated to health care in the political economy extension, in this section we carry out numerical experiments to gain insights into political economy outcomes and the long run behaviour of the model, particularly in relation to the wealth dynamics created through the political economy process. We do not report the results of additional experiments carried out by reducing the variance of the initial wealth distribution and varying other parameters in the model, as these changes do not affect the general pattern of the results. We emphasise that this section is not intended to be a comprehensive calibration exercise; we only wish to illustrate some qualitative features of the model using a reasonably plausible (rather than exact) parameterisation of the model.[40]

For the experiments reported in this section, we assume that the economy is populated by 501 agents and that the initial wealth distribution is lognormal with mean 4.2 and standard deviation 0.8, which results in an initial sample-mean wealth of 91.0988 and a Gini coefficient of 0.4259. In addition, we assume the following parameter values: $x = 0.03$, $\varepsilon = 0.8$, $\bar{\phi} = 0.96$, $\tau = 0.225$, $\sigma = 0.88$, $\bar{r} = 0.3217$, $\theta = \left(\frac{1}{\bar{r}}\right)^{\sigma} + 0.01$, $\bar{w}_1 = 23$ and $\bar{w}_2 = 0.2\bar{w}_1$.[41] The values of x and ε are chosen such

[40] As we have conducted comprehensive sensitivity analyses for various parameters, we are confident that the results presented below are indeed representative of the qualitative features of the model. We do not present these here; however, they are available upon request.

[41] Although Chakraborty and Das (2005) assume that the agent earns a wage of $\bar{w}$ in each period, we assume that the agent earns a lower wage in old age. While this assumption is essential to prevent the emergence of negative bequests in our model, it is also reasonable in light of empirical evidence. For instance, according to Fernández-Ballesteros et al. (2011), older workers are mostly involved in volunteering and helping family members, and rewards for such activities are typically very low or non-existent. Furthermore, Verhaeghen and Salthouse (1997) reveal through a meta-analysis incorporating 91 studies that speed, reasoning ability, spatial ability, and memory generally deteriorates sharply after the age of 50. Skirbekk (2004) shows that there is a decrease in productivity with age in jobs where the main skills required are problem solving, the ability to learn and performing activities speedily, but older workers can still be productive in jobs that demand verbal skills and experience.

that the minimum survival probability resulting from the experiments is 0.3043.[42] The top marginal tax rate levied on gifts and bequests earned by children varies between 0% and 45% in the OECD (OECD, 2013, p. 80) and the value of τ used in these experiments is the mid-point of these two values. Given that in this model, endogenous time preference entails $\sigma < 1$, and further noting the suggestion by Prescott (1986) that the value of σ "cannot be far from 1", we use $\sigma = 0.88$ in our experiments. Taking an agent's average lifetime to be 80 years, if we assume that decision making commences at the age of 20, each period in the model corresponds to 30 years. Therefore, we take the interest rate $\bar{r}$ to be equal to the OECD average return on savings after 30 years, which is approximately 32.17% (OECD, 2013, p. 78). The value of $\bar{\phi}$ is roughly equal to 1 minus the 50-59 mortality rate in the United Kingdom in 2010 (Whitlock, 2012). We carry out experiments on three values of a: 0.8, 0.5 and 0.3 and the values of b we consider are: -0.8, -0.3, 0, 0.2, 0.7 and 1.2.[43]

5.5.1 Wealth dynamics

In this section we examine how average levels and inequality of wealth in the political economy model evolve over time. Generally, we observe that a higher value of the aggregate substitutability parameter b is associated with a higher average wealth level and lower inequality in the long run. Furthermore, as suggested in Proposition 2, the long run wealth distributions resulting from all experiments are bimodal.

---

[42] This survival probability is very close to 1 minus the average adult mortality rate in the African region (IHME, 2014). As this is a model with heterogeneous agents, we calibrate the lower end of the spectrum of mortality based on developing country data, while the maximum survival probability is defined from developed country data.

[43] In Section 2, we demonstrated that $b < \frac{1}{a}$ is necessary to prevent a health capital and survival probability of 0. The values of b chosen for the experiments adhere to this upper bound across all three values of a considered.

5.5.1.1 Long run average wealth

The experiments reveal that a higher value of the aggregate substitutability parameter b generally yields a higher long run wealth level.[44] This is illustrated in Figure 5.4, which shows the evolution of long run wealth for different values of b given $a = 0.8$. An interesting observation relating to the result that a higher degree of aggregate substitutability between public and private health inputs is associated with higher long run wealth levels, is that it bears a resemblance to the positive association between long run income per capita and the elasticity of substitution between labour and capital in the production function proposed by authors such as Klump and De La Grandville (2000), Klump and Preissler (2000); and Palivos & Karagiannis (2010).[45] Despite the similariy of outcomes, such a comparison needs to be viewed with caution, as the results of this study refer to the elasticity of substitution between public and private health expenditures, which, as noted previously, has a bearing upon preferences and is conditioned by factors quite different to those that determine the elasticity of substitution between capital and labour.

[44] For a small range of values of the parameter a, the relationship between average long run wealth and b is mildly non-monotonic. For the values of a considered in the experiments reported in this section, this occurs when $a = 0.3$. However, even in this instance, we observe a high correlation of 0.9 between the values of b and average wealth levels.

[45] For instance, Klump and De La Grandville (2000) demonstrate that in a Solow-Swan model with a normalised CES production function, a country with a higher elasticity of substitution between capital and labour is able to grow at a faster rate and also enjoy a higher income and capital stock per capita at steady state.

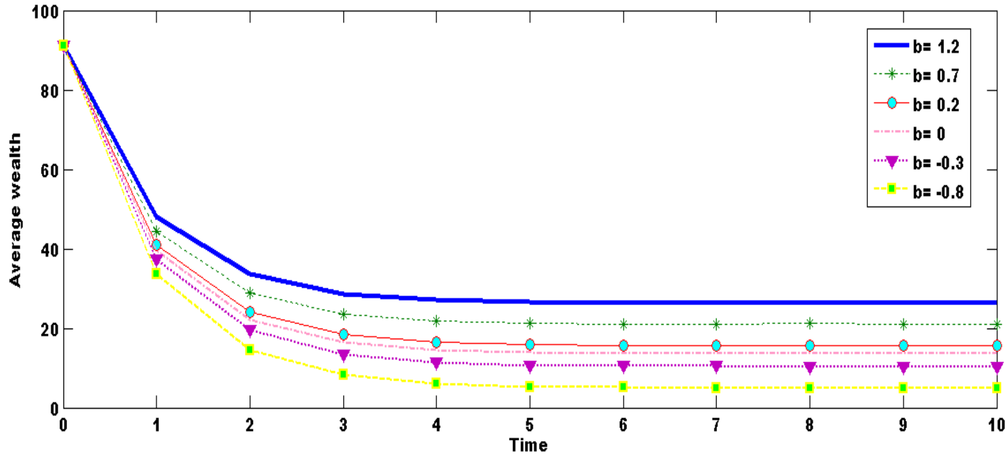


Figure 5.4: Behaviour of average wealth over time for different values of b given $a = 0.8$

In addition to the value of b, long run average wealth is also affected by the value of a, which we interpret as the "pure" share of public spending in the health production function. Figure 5.5, which plots the long run wealth against b for different values of a, shows that in general, a higher value of a is associated a higher long run average wealth level. The intuition for this result is rather straightforward, in that a higher value of the pure share of public spending a implies that public expenditure can have a bigger impact on a person's health capital, which enables an individual to achieve a given level of health capital by incurring less private health expenditure. Lower private health expenditure allows an individual to leave higher intended or unintended bequests for progeny, thereby yielding higher long run wealth levels.

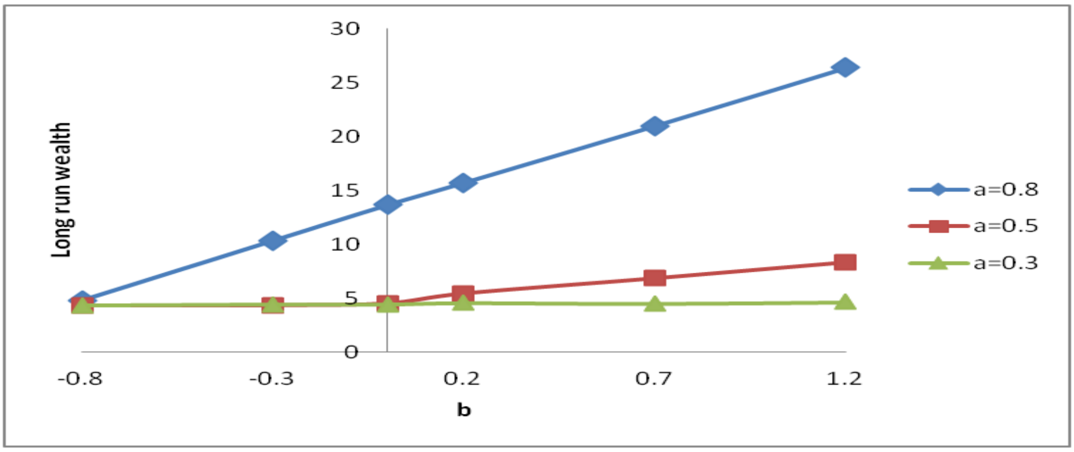


Figure 5.5: Long run average wealth for different values of a and b

5.5.1.2 *Long run wealth inequality*

In addition to exploring the manner in which the degree of aggregate substitutability and the pure share of public spending influence the long run average levels of wealth, it is also of interest to examine how they affect the long run inequality of wealth where inequality is measured by the Gini coefficient.[46] While inequality tends to decrease over time, which one would expect, since there is redistribution through public spending on health in this economy, a higher value of the aggregate substitutability parameter b is generally associated with a lower level of inequality. Thus, a higher value of b, in addition to yielding a higher average level of wealth, can also lead to lower inequality in the long run. This is illustrated in Figure 5.6, which shows the evolution of the Gini coefficient over time when $a = 0.8$. Although a clear negative association between the aggregate substitutability parameter b and long run wealth inequality is present for a large range of values of a for the relatively lower

[46] We measure inequality using the formula for the Gini coefficient supplied in World Bank (2011) which is: $G = \dfrac{2Cov(Y_i, \overline{F_i})}{\overline{Y}}$, where G : Gini coefficient, Y_i income/wealth of household i , $\overline{Y}$: mean income/wealth and F : rank of household *i* (0 for poorest, 1 for richest).

value of $a = 0.3$, the relationship between b and long run inequality is non-monotonic, with a correlation coefficient of -0.1469 between the Gini coefficient and b when $a = 0.3$, suggesting the presence of a weak negative linear association between the two variables.

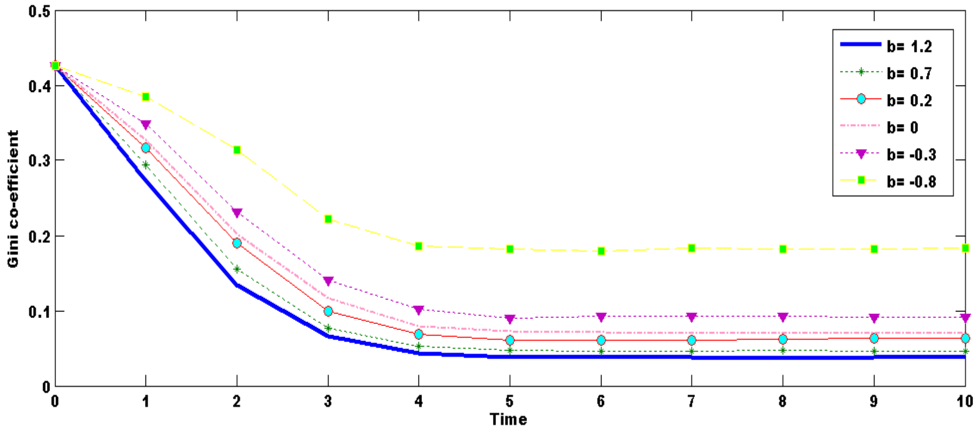


Figure 5.6: Behaviour of the Gini coefficient over time for different values of b, given $a = 0.8$

Figure 5.7 illustrates that a higher value of the pure share of public spending a is also generally associated with a lower level of inequality. This is created by the obvious tension in the production function created by the parameter a. Being the exponent of public health spending, a higher value of a provides a bigger boost to the efficiency of public health spending whilst reducing the exponent of the combined public and private term, and its efficiency. Hence, a higher value of a reduces the value of private health expenditures agents undertake to complement public healthcare, and this, in turn, augements the redistributive role of public health care spending by enabling even the poorer agents in the economy to leave more bequests for progeny. Hence, higher values of the aggregate substitutability parameter b and the pure share of public spending a have a beneficial impact upon the economy, as they result in higher average wealth levels and lower wealth inequality in the long run. However, the non-monotonicity we observe when $a = 0.3$ suggests that the beneficial outcomes associated with a higher aggregate substitutability parameter are dependent upon the pure share of public spending in the health production function being sufficiently high.

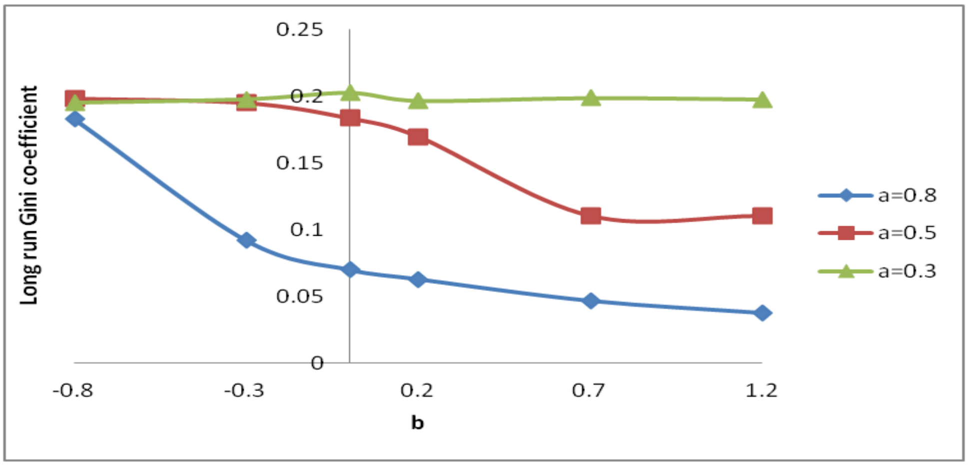


Figure 5.7: Long run Gini coefficient for different values of a and b

5.5.2 Political economy versus socially optimal results

It is often the case that the political economy outcome is not 'optimal' relative to a policy that maximises the welfare of agents in the economy, where social welfare can be measured in different ways. Hence, in this section, we explore the behaviour of the winning value of the ψ, which is the proportion of tax revenue allocated towards public health care over time for the different values of the aggregate subsitutability parameter b and the pure share of public spending a. We also compare these winning values of ψ against the welfare maximising values of ψ under the utilitarian and Rawlsian approaches.[47] For a large range of parameter values, the voting outcomes are majoritarian, and indeed, in large number of instances, they are unanimous.[48] In most

[47] The utilitarian or Benthamite welfare function maximises the sum of utilities of all agents in the economy, while under the Rawlsian or maxi-min criterion, the utility of the least well off member/s of society is maximised. These two approaches represent two different ideologies to the maximisation of welfare, with the first maximising the sum of utilities and the second maximising the utility of the worst-off individual (Ng, 2003, pp 128).

[48] In instances where a winning value cannot be found through the majority rule, the plurality rule is employed, where the outcome with the highest number of votes is considered the winning one.

cases, the winning values correspond to the welfare maximising values in the long run, although there are a few instances where the Rawlsian and utilitarian welfare maximising values differ from one another and the political economy result converges to one of them in the long run.

When $a = 0.8,$ numerical experiments reveal that there is a unaminous vote in favour of a proportion of ψ equal to 1 in all periods for all values of the aggregate substitutability parameter and that the welfare maximising value of ψ under both the utilitarian and Rawlsian approaches is also equal to 1. When $a = 0.5,$ for $b \geq 0,$ all agents vote unanimously for a value of ψ equal to 1 in all periods and this winning value is consistent with welfare maximisation under both the utilitarian and Rawlsian approaches in all periods. With regard to the negative values of b considered when $a = 0.5,$ for $b = -0.8$, the winning and the welfare maximising values values converge to 0.75 in the long run. However, when $b = -0.3,$ the winning value of ψ converges to the Rawlsian welfare maximising value of 0.9, but the welfare maximising value under the utilitarian approach converges to 0.95. Hence, when $a = 0.5,$ we see a tendency for the winning value of the proportion of tax revenue allocated to health care to increase with the value of the aggregate substitutability parameter b.

This positive relationship between the winning value of ψ and the aggregate substitutability parameter b is even more prominent when $a = 0.3;$ with the winning value of value of ψ gradually changing from 0.5 when $b = -0.8$ to 1 when $b = 1.2.$ The typical behaviour of the winning and welfare maximising values of ψ over time is illustrated in Figure 5.8, where the path of the winning value is compared to the paths of the utilitarian and Rawlsian welfare maximising values when $b = 0.7.$ Figure 5.8 shows that in the transition path, the winning value of ψ falls between the the utilitarian and Rawlsian outcomes. Recall that the remaining proportion $(1-\psi)$ of tax revenue is distributed among individuals as a the lump-sum transfer. Given that this lump sum transfer is a perfect substitute for consumption, it is natural that the poorer agents in the economy would benefit more from this lump sum transfer. This results in the Rawlsian welfare maximising value being relatively lower than that obtained under the utilitarian criterion. It is also important to note that public health care is characterised by an externality, in that it impacts the survival probability of all individuals in the economy positively. The utilitarian approach, which weighs the utilities of all individuals equally, is able to capture this externality, thereby contributing towards internalising it. This provides an explanation for the comparatively higher social welfare maximising value obtained under the utilitarian criterion. As the individual takes into account only her own welfare when voting on the value of ψ, the political economy outcome does not capture the externality effect, nor does it attach any particular importance to the utility of the poorest members of society, and one would therefore expect it to lie between the Rawlsian and utilitarian values.

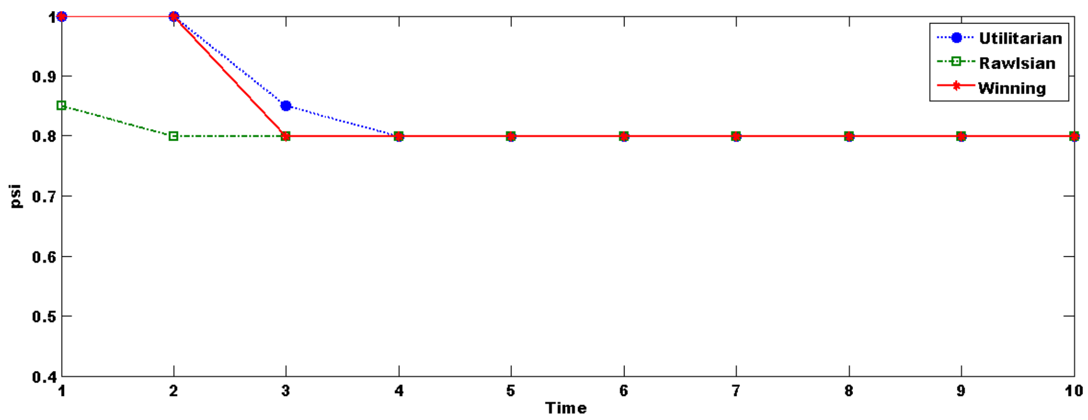


Figure 5.8: Behaviour of the winning and welfare maximising values of ψ over time for $b = 0.7$ given $a = 0.3$

It is also worth noting that in the interim periods, the winning value is usually closer to the utilitarian welfare maximising value rather than the Rawlsian value. An explanation for this is the similarity between the utilitarian approach, which weighs each individual's utility equally, and the political economy outcome, which is derived by weighing each agent's vote equally, where the vote depends on the agent's utility. The fact that the political economy outcome is closer to the utilitarian outcome also suggests the possibility that in the transition path, the middle and the upper classes, who desire greater public spending on health care, combine to defeat the poor agents in the economy who prefer a comparatively lower proportion of tax revenue to be allocated towards health care. Consequently, the resulting voting outcomes in the transition path, although sub-optimal from a social welfare maximising point of view, are relatively more favourable towards the middle and the upper classes rather than the poorest agents in the economy. This resulting outcome is quite different from the "ends against the middle" result presented in Epple and Romano (1996a), who show that in the presence of private alternatives to a publicly provided good, middle-income households will vote for increased public provision, as opposed to the rich and poor households who prefer reduced private provision. Recall that in our model, the government's tax revenue is divided between public health spending and a lump sum transfer. A possible reason for the emergence of this "middle and upper versus poor" scenario is the higher probability of premature death of the poor agents. As a result of this, they prefer the lump-sum transfer and the associated increase in consumption it can bring about over public expenditure on health care.

Although the winning and the welfare maximising values converge in the long run for most parameter values considered, in a few instances, the long run utilitarian and Rawlsian welfare maximising values of ψ differ from each other, and the winning value of ψ converges to one of them. One such case occurs when $b = 0.2$ given $a = 0.5$. In this instance, the winning value converges to the Rawlsian value. We observe such non-convergence when $b = -0.3$ given $a = 0.3$ as well, but as illustrated in Figure 5.9, the long run political outcome converges to the utilitarian value rather than the Rawlsian one in this instance. The other instance with non-convergence that occurs when $a = 0.3$, also illustrated in Figure 5.9, is when $b = 0.2$, where the winning and Rawlsian welfare maximising value converges to 0.65, while the long run utilitarian value is 0.7. It is difficult to provide an intuition for the behaviour depicted in these instances, especially the "cross-over" we observe in the case of $b = 0.2$. However, in all these cases, the differences in the long run Rawlsian and utilitarian values are very small.

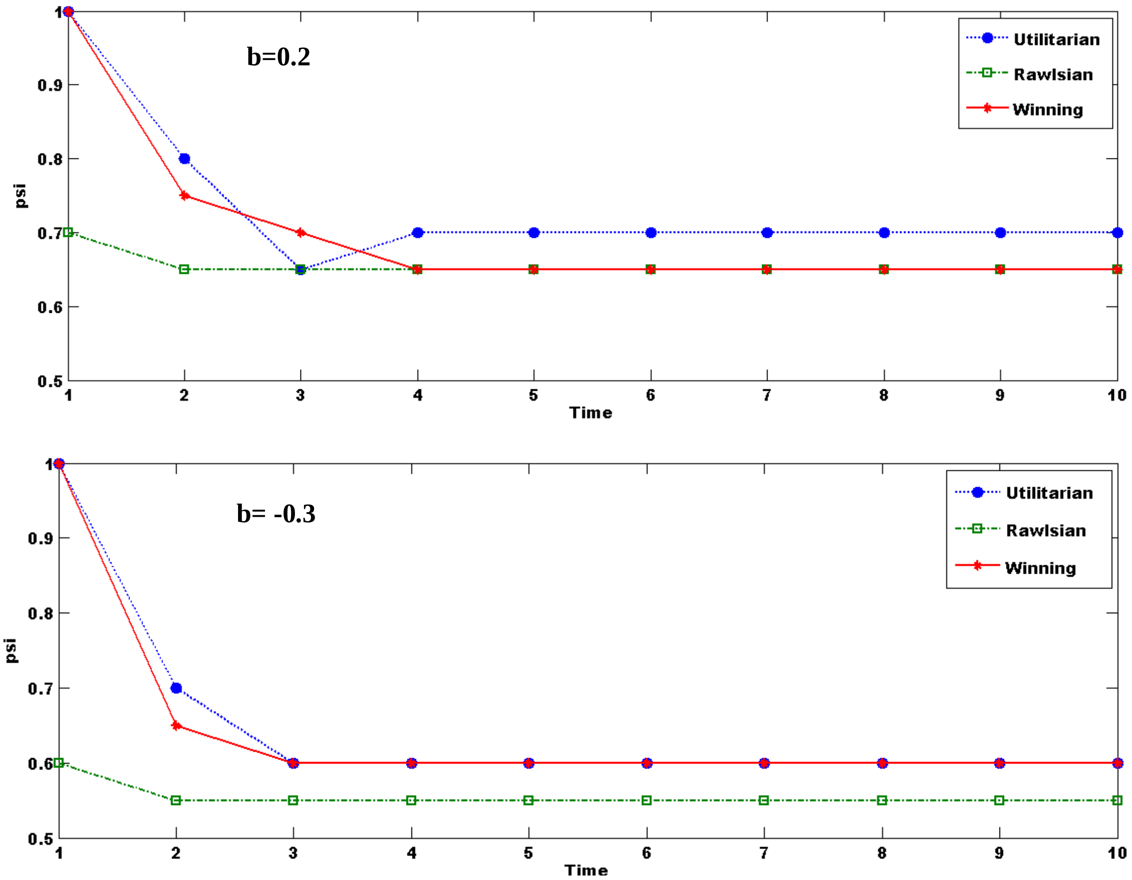


Figure 5.9: Behaviour of the winning and welfare maximising values of ψ over time for $b = 0.2$ and $b = -0.3$ given $a = 0.3$

The tendency for the winning value of ψ to increase with the aggregate substitutability parameter b is consistent with Proposition 4, which sugggests that the preferred value of ψ is increasing in b for all agents. The fact that the preferred value of ψ for all agents in the economy is positively related to the aggregate susbtitutability parameter b manifests itself through the political economy outcome. One would expect the aggregate substitutability parameter b to be high in developed countries where health services offered by both the public and private sectors are of a comparable quality. However, in developing countries, the degree of aggregate substitutability is likely to be lower, as private health services tend to be of a superior quality to public health care. Hence, the results suggest that people in developed countries, where aggregate substitutability is higher, prefer a higher proportion of tax revenue to be allocated towards the provision of public health care when compared to citizens of developing countries.[49]

Indirect empirical support for the positive association between the aggregate substitutability parameter b and the proportion of tax revenue allocated towards public health care is provided by studies such as Hardie and Critchley (2008), Soroka (2007) and Wilson and Rosenberg (2004), which demonstrate the presence of strong public support for continued public provision of health care services in some of the richest countries in the world, such as Australia and Canada. Indeed, in the EU as a whole,

[49] Studies have revealed that in some countries, private hospitals are more efficient that public ones. For instance, a study by Masiye (2007) reveals that in Zambia, government owned hospitals have a lower mean efficiency score when compared to private and mission owned hospitals. In the case of Taiwan too, H. Chang, Cheng, and Das (2004) use non-parametric data envelopment analysis to show that private hospitals outperform public ones. When private hospitals are comparatively more efficient than public ones, the aggregate substitutability between public and private health expenditures is likely to be lower.

public health expenditure accounts for 14.9% of total government expenditure and represents a massive 7.3% of the region's GDP (European Commission, 2014), suggesting that people in developed countries prefer a high degree of state involvement in the funding of health care. In contrast, in a number of African countries, public health spending accounts for less than 5% of total government expenditure, and in many of them, public health care spending per capita is well below the minimum of 34 US dollars recommended by the World Health Organisation to supply people with basic health care facilities (KPMG, 2012).

As discussed earlier, the notion of substitutability in our model has a connection with the relative quality of public versus private health care. In that sense, the results of our model provide a potential explanation for the above mentioned differences in public health spending between developed and developing countries. Essentially, the model provides a political economy argument for the small degree of state involvement seen in developing countries. Agents in these economies do not vote for a high proportion of tax revenue to be allocated towards public health care, as it is a relatively ineffective substitute for private health services.[50]

While we observe a positive relationship between the value of the aggregate substitutability parameter b, we also notice that the winning value of ψ is increasing in the value of a, which we interpret as the pure share of public spending in the health production function, This relationship is illustrated in Figure 5.10, which shows the winning values of ψ against values of b for different values of a. Recall that the preferred proportion of tax revenue to be allocated towards public health care for

[50] However, it is also important to appreciate that in reality, low public spending on health care in developing countries may not occur due to voting outcomes that occur within a democratic setting, but could instead be the result of decisions made by corrupt and authoritarian governments.

agents below and above the critical wealth level $\hat{W}_t$ was given by (5.3.14) and (5.3.15). If one observes these values, it is clear that they are increasing in the parameter a, which leads to the observed positive association between the pure share of public spending a and the winning value of the proportion of tax revenue to be allocated towards health care ψ. The rationale for this outcome is that a higher value of a means that public health spending can make a greater impact upon a person's health capital and thereby survival probability, it then follows that agents would wish for a high proportion of government revenue to be dedicated towards the provision of health care.

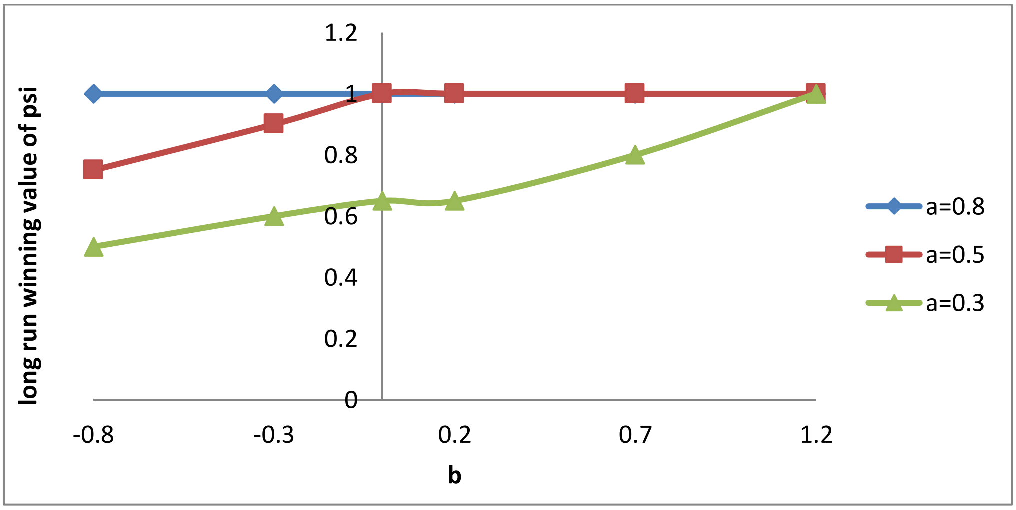


Figure 5.10: Long run winning value of ψ against b for different values of a

Across all of the experiments considered, we observe that the winning value of ψ ranges between a minimum of 0.5 and a maximum of 1, suggesting that a sizeable proportion of tax revenue is allocated towards health care in this economy. This result may be rather "artificial" because in our health production function, public health care is the essential input and this means that voting for zero public funding of health care is not an optimal decision for any agent in this model. A value of $\psi = 0$ would cause the level of health capital to collapse to 0, thereby yielding a survival probability and utility of 0. In the VES construct, there are only two possibilities—either public or private health can be considered as the essential input. Our choice is motivated by the considerations outlined earlier about the relative importance of the public health care system.

5.6 CONCLUSION

This paper examines the impact of the aggregate degree of substitutability between public and private health expenditures on long run macroeconomic outcomes using an overlapping generations model with heterogeneous agents, in which the probability of survival into old age depends on health capital accumulated in youth The novelty of this paper lies in the use of a variable elasticity of substitution "health production function", in which a person's health capital is determined by public and private health expenditures, and each individual's degree of substitutability between public and private health expenditures is influenced by a parameter b, which interpret as the degree of aggregate substitutability, as well as her wealth. Hence, with heterogeneous wealth endowments, this functional form results in the degree of substitutability between public and private health inputs to differ between individuals.

We have shown that a higher value of the aggregate substitutability parameter b leads to higher average wealth levels, as well as lower wealth inequality in the long run. We also demonstrate that a higher value of another parameter a, which we refer to as the "pure" share of public spending in the health production function, is also generally associated with a higher average level and lower inequality of wealth in the long run. Political economy outcomes generally reveal that agents in this economy vote for a higher proportion of tax revenue to be allocated towards public health care spending when the degree of aggregate substitutability and the pure share of public spending in the health production function are higher, and that in most cases, the winning values are consistent with social welfare maximisation. As aggregate substitutability is likely to be higher in developed countries where public and private health services are of a comparable quality, the fact that public spending is relatively much higher in these countries when compared to developing ones lends empirical support for this result.

The results of the paper carry several policy implications. As discussed above, a greater degree of substitutability and a larger share of public health expenditures in determining an individual's health capital yield better long run macroeconomic outcomes. Therefore, policies aimed at improving institutional aspects that improve the degree of substitutability between public and private health inputs are likely to have beneficial effects on economic growth and development. In many developed countries, there is a large state sector that generally provides high quality health care services to the population. Usually, a parallel private sector providing comparable services exists in many of these countries (Atella & Deb, 2008; Cheng & Vahid, 2010), leading to a high degree of substitutability between public and private health inputs at the aggregate level. However, in some developed countries, the private sector is focussed on serving small niches of the market (Tuohy, Flood, & Stabile, 2004), may face various regulations (Lülfesmann & Myers, 2010), and is relatively more expensive when compared to public health care, which is usually free or is covered by state insurance (Fabbri & Monfardini, 2009). These factors restrict the degree of substitutability between public and private health care. Therefore, strengthening private finance and deregulating the health sector, which would improve availability and affordability of private health services, would help raise the degree of aggregate substitutability between public and private health services in these countries.

In the case of less developed countries, a low degree of aggregate substitutability may stem from low service quality and poor facilities in the public sector that encourage people to opt for private health services that are of a comparatively higher quality. Therefore, in the context of a developing country, making public health care a viable competitor to private health care through better funding, service quality, and geographical coverage is a possible policy option to improve the aggregate substitutability between public and private health services.

Given that a higher "pure" share of public spending in the health production function also has a positive impact on long run outcomes, policies aimed at improving this dimension could also be considered. The pure share of public spending in the health production function captures the ability of public spending to influence an individual's health capital, and could therefore be regarded as a measure of the effectiveness of public health care. Thus, possible policies to improve the pure contribution of public health care could include technological improvements in the public health care system, dedicating more resources towards training public health care workers in order to raise their productivity, a greater focus on state-funded medical research, and investment in new equipment.

The results of the paper are also amenable to empirical research. Given the important influence of aggregate substitutability on long run macroeconomic outcomes revealed in the study, a direction for future research suggested by the paper is the possibility of developing cross-country measures of aggregate substitutability by using aspects such as the relative quality of public health care and the regulatory framework in the health care sector as proxy variables.

Chapter 6: Concluding Remarks

This thesis provides some diverse insights into several macroeconomic issues such as growth patterns, wealth inequality, policy, and welfare using dynamic general equilibrium models. A brief summary of the key results of the three studies that constitute this thesis and discussion of the policy recommendations and other implications emanating from them conclude the thesis. The limitations of the studies and the directions for future research emerging from the outcomes of these studies are also briefly discussed.

At the heart of the three essays constituting this thesis is the variable elasticity of substitution (VES) production function, a form of production function originally discussed by Sato and Hoffman (1968) and Revankar (1971). In the VES form, the elasticity of factor substitution is an *endogenous variable*. Although the growth literature often assumes that production follows a CES or Cobb-Douglas technology, there is substantial empirical evidence that has raised doubts about the primary assumption underlying these forms, which is that the elasticity of factor substitution remains constant across input combinations and over time. In light of such observations, the VES production function is an intuitively appealing alternative to these production functions. In the VES form, the elasticity of factor substitution is linearly related to the capital stock per worker through a technology parameter that we refer to as b. Endogeneity of the elasticity of factor substitution in the VES from stems from the fact that the capital stock per worker is closely associated with the rate of economic growth.

In the first essay, we explore the properties of the VES production function and apply it in the context of an otherwise standard Diamond overlapping generations model. The exploration of the properties of this form yields several interesting results. Among them are the following: a higher value of the technology parameter is associated with a higher return to capital, a more capital intensive long run expansion path, as well as a declining share of labour in total output. The parameter b also affects the direction of technical change in the economy, as a higher value of it causes greater capital bias in production. When the VES form is applied to a Diamond model, we observe that there could be several possible long run growth trajectories. The economy could converge to a unique and stable steady state akin to that in the standard Solow-Swan model, regress towards a poverty trap, or diverge towards an upper limit of capital per worker hinging upon the parameter restrictions associated with the VES form.

There is potential to strengthen these findings through further empirical work. Empirical work on the VES production function has emerged only sporaidically up to now, and in our view, they have not deployed the inherent flexibility offered by the construct effectively. For instance, Karagiannis et al (2005) use cross-country panel data to estimate a single value for the elasticity of factor substitution using the VES production function. However, by deploying the variable nature of the VES form, one could obtain different estimates of the technology parameter b, and hence, the elasticity of substitution for different countries, as well as explore how the elasticity of factor substitution has evolved over time within countries. From the point of view of the results of this thesis, as well as recent empirical growth literature, such a study can shed light on the mechanics behind directed technical change. This exploration is a potentially attractive offshoot of the work presented in the first essay. Furthermore, a comparison of the growth dynamics emerging from different frameworks, possibly the Solow, Diamond, and Cass-Coopmans models is also an interesting theoretical extension to the first study.

The public-private mix in the provision of public goods such as health and education is an issue that has attracted considerable attention in the economics profession in recent decades. Given that the individual plays the primary role in the creation of her own human capital, we assume that she combines public and private inputs in health care or education to "produce" these goods. The degree to which these inputs are substitutable for one another can be a key determinant of long run macroeconomic outcomes. Furthermore, government spending on public health care and education is typically determined through a voting mechanism. Hence, the degree to which public and private inputs are considered by individuals to be substitutable to one another is likely to be an important determinant of an economy's political economy outcomes. The flexibility of the VES form and its associated properties makes it an appropriate construct to employ in the exploration of such issues. To this end, in the second and third studies, we apply the VES form to study the macroeconomic ramifications emerging from the degree of substitutability between public and private inputs in the contexts of education and health respectively. In the context of such human capital production functions, b can be interpreted as a substitutability parameter that captures institutional and cultural aspects that affect the combination of public and private inputs that the individual selects.

In the second study, we develop an overlapping generations model with a VES-type education production function with public and private inputs. Education acquired in youth, coupled with parental human capital, are the determinants of an agent's adult age human capital, which she combines with physical capital to produce output. This framwork yields a number of possible dynamics for the evolution of human and physical capital in the economy. We observe that the economy could display oscillatory or monotonic convergence towards a stable steady state or display oscillatory divergence followed by permanent cycles. Optimal policy considerations point in the direction that a higher value of the parameter b is associated with a higher tax rate to finance public education.

Along a similar vein, the third essay looks at these considerations in the context of health. However, the fundamental difference in this study is that it considers a heterogenous agents framework where agents have different initial endowments of wealth. The health capital the agent acquires is primarily determined by the wealth she inherits from her parent. This health capital, in turn, is the key determinant of her probability of survival into old age. Unlike in the case of a homogenous agents framework, such as that considered in the second study, in the case of the heterogenous agents framework developed in this study, the VES form enables a distinction to be made between substitutability between public and private health inputs at the aggregate level, captured by the parameter b, as well as individual substitutability, represented by the elasticity of substitution value corresponding to each individual, which hinges on the ratio of public to private inputs she selects.

The analytical and numerical results reveal that a higher value of the parameter b is associated with higher average wealth levels, as well as lower inequality in the long run. In terms of political economy considerations, we observe that higher the value of this parameter, agents vote for a higher rate of wealth tax to finance public health care.

We argue that the level of aggregate substitutability between public and private inputs in the case of both health care and education is likely to be higher in developed economies, due to both sectors providing services of comparable quality. Hence, the results of the second and third studies provide a potential, albeit indirect explanation for the higher government expenditures on health and education in these countries.

With regard to the directions for further research emerging from the second and third studies, the estimation of the hypothesised VES health and education production functions is one possible offshoot of the work presented herein. Specifically, an active possibility would be to employ productivity and efficiency analysis to estimate the efficient "frontier" using the VES production functions used in these studies and compare efficiencies across and within countries. Given that much of the extant research on productivity and efficiency analysis uses the CES production function, the use of the VES form would be an innovative contribution in itself. Possible theoretical extensions include the use of different frameworks, for instance, a two sector model, and the combination of health and education considerations into a single framework.

The thesis contains several limitations. The first of these is the VES form itself; the boundededness of capital per worker from above results in the inability of the form to create unbounded growth, thereby limiting the range of trajectories the VES form can capture. Furthermore, the strict linearity of the elasiticity of substitution with respect to the capital stock per worker (or in the context of the second and third studies, the ratio of public to private inputs) is also considered to be another problem inherent to the VES form, as it limits its flexibility, although in the contexts of education and health considered in Chapters 4 and 5, this feature actually improves, rather than diminishes, flexibility. Neverthless, these limitations call for further refinement of the VES form.

Another problem, particularly in the case of the second and third essays, is the fact that "aggregate substitutability" is not directly measurable. Given that the bulk of these results hinge on this parameter, there is a need to quantify this construct in practical terms. Hence, we hope that the theoretical results of these studies will serve as a basis for further, more specific efforts to quantify this notion.

Overall, this thesis makes some interesting contributions to several spheres in the area of dynamic macroeconomics. However, as noted earlier, it is possible to develop many possible extensions and offshoots, both theoretical and empirical, based on the work presented herein. We hope that our work will provide a motivating foundation for further research in this area.

Bibliography

Abdih, Y., Chami, R., Dagher, J., & Montiel, P. (2012). Remittances and institutions: Are remittances a curse? *World Development, 40*(4), 657-666.

Abel, A. B. (1985). Precautionary saving and accidental bequests. *The American Economic Review, 75*(4), 777-791.

Abramovitz, M. (1990). The catch-up factor in postwar economic growth. *Economic Inquiry, 28*(1), 1-18. doi:10.1111/j.1465-7295.1990.tb00800.x

Acemoglu, D., Johnson, S., & Robinson, J. A. (2000). *The colonial origins of comparative development: An empirical investigation* (No. w7771). National Bureau of Economic Research.

Acemoglu, D. (2002). Directed technical change. *The Review of Economic Studies, 69*(4), 781-809.

Acemoglu, D., & Robinson, J. A. (2002). The political economy of the Kuznets Curve. *Review of Development Economics, 6*(2), 183-203. doi:10.1111/1467-9361.00149

Acemoglu, D., & Ventura, J. (2002). The world income distribution. *The Quarterly Journal of Economics, 117*(2), 659-694. doi:10.1162/003355302753650355

Acemoglu, D. (2003). Labor- and capital-augmenting technical change. *Journal of the European Economic Association, 1*(1), 1-37. doi:10.1162/154247603322256756

Acemoglu, D., Johnson, S., & Robinson, J. A. (2005). Institutions as a fundamental cause of long-run growth. *Handbook of economic growth, 1,* 385-472.

Acemoglu, D. (2008). *Introduction to modern economic growth*: Princeton University Press.

Aggarwal, R., Demirgüç-Kunt, A., & Pería, M. S. M. (2011). Do remittances promote financial development? *Journal of Development Economics, 96*(2), 255-264.

Aghion, P., & Howitt, P. (2009). *The Economics of Growth* (Vol. 1): The MIT Press.

Akhmedjonov, A., Lau, M. C. K., & İzgi, B. B. (2013). New evidence of regional income divergence in post-reform Russia. *Applied Economics, 45*(18), 2675-2682.

Alesina, A., & Perotti, R. (1996). Income distribution, political instability, and investment. *European Economic Review, 40*(6), 1203-1228. doi:10.1016/0014-2921(95)00030-5

Alesina, A., & Rodrik, D. (1994). Distributive politics and economic growth. *The Quarterly Journal of Economics, 109*(2), 465-490. doi:10.2307/2118470

Alexander, K. L., & Pallas, A. M. (1983). Private schools and public policy: new evidence on cognitive achievement in public and private schools. *Sociology of Education, 56*(4), 170-182.

Arcalean, C., & Schiopu, I. (2010). Public versus private investment and growth in a hierarchical education system. *Journal of Economic Dynamics & Control, 34*(4), 604-622. doi:10.1016/j.jedc.2009.11.006

Arrow, K. J., Chenery, H. B., Minhas, B. S., & Solow, R. M. (1961). Capital-labor substitution and economic efficiency. *The Review of Economics and Statistics, 43*(3), 225-250.

Atella, V., & Deb, P. (2008). Are primary care physicians, public and private sector specialists substitutes or complements? Evidence from a simultaneous equations model for count data. *Journal of Health Economics, 27*(3), 770-785. doi:10.1016/j.jhealeco.2007.10.006

Baier, S. L., Dwyer, G. P., & Tamura, R. (2006). How important are capital and total factor productivity for economic growth? *Economic Inquiry, 44*(1), 23-49. doi:10.1093/ei/cbj003

Bairam, E. (1989). Learning-by-doing, variable elasticity of substitution and economic growth in Japan, 1878-1939. *Journal of Development Studies, 25*(3), 344-353. doi:10.1080/00220388908422116

Baldwin, R. E., Martin, P., & Ottaviano, G. I. (2001). Global income divergence, trade, and industrialization: The geography of growth take-offs. *Journal of Economic Growth, 6*(1), 5-37.

Barajas, A., Chami, R., Fullenkamp, C., Gapen, M., & Montiel, P. J. (2009). Do workers' remittances promote economic growth? *IMF Working Papers*, 1-22.

Barro, R. J., & Sala-i-Martin, X. (1992). Convergence. *Journal of Political Economy, 100*(2), 223-251. doi:10.2307/2138606

Bearse, P., Glomm, G., & Patterson, D. M. (2005). Endogenous public expenditures on education. *Journal of Public Economic Theory, 7*(4), 561-577. doi:10.1111/j.1467-9779.2005.00234.x

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy, 70*(5), 9-49. doi:10.2307/1829103

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press.

Becker, G. S. (1994). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. US: University of Chicago Press.

Becker, G. S. (2007). Health as human capital: synthesis and extensions. *Oxford Economic Papers, 59*(3), 379-410.

Becker, G. S. (2009). *A treatise on the family*: Harvard University Press.

Beine, M., Docquier, F., & Oden-Defoort, C. (2011). A panel data analysis of the brain gain. *World Development*, *39*(4), 523-532. doi:10.1016/j.worlddev.2010.03.009

Beine, M., Docquier, F., & Rapoport, H. (2001). Brain drain and economic growth: theory and evidence. *Journal of Development Economics*, *64*(1), 275-289. doi:10.1016/S0304-3878(00)00133-4

Beine, M., Docquier, F., & Rapoport, H. (2008). Brain drain and human capital formation in developing countries: Winners and losers. *The Economic Journal, 118*(528), 631-652. doi:10.1111/j.1468-0297.2008.02135.x

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy, 75*(4), 352-365. doi:10.1086/259291

Bénassy, J. P., & Brezis, E. S. (2013). Brain drain and development traps. *Journal of Development Economics, 102*, 15-22. doi:http://dx.doi.org/10.1016/j.jdeveco.2012.11.002

Benhabib, J. J., & Spiegel, M. M. (1994). The role of human capital in economic development evidence from aggregate cross-country data. *Journal of Monetary Economics, 34*(2), 143-173.

Beraldo, S., Montolio, D., & Turati, G. (2009). Healthy, educated and wealthy: A primer on the impact of public and private welfare expenditures on economic growth. *The Journal of Socio-Economics, 38*(6), 946-956.

Bertola, G. (1996). Factor shares in OLG models of growth. *European Economic Review, 40*(8), 1541-1560. doi:http://dx.doi.org/10.1016/0014-2921(94)00107-3

Bhagwati, J., & Hamada, K. (1974). The brain drain, international integration of markets for professionals and unemployment. *Journal of Development Economics, 1*(1), 19-42. doi:10.1016/0304-3878(74)90020-0

Bhattacharya, J., & Qiao, X. (2007). Public and private expenditures on health in a growth model. *Journal of Economic Dynamics and Control, 31*(8), 2519-2535. doi:10.1016/j.jedc.2006.07.007

Bidani, B., & Ravallion, M. (1997). Decomposing social indicators using distributional data. *Journal of Econometrics, 77*(1), 125-139. doi:http://dx.doi.org/10.1016/S0304-4076(95)01809-3

Birdsall, N., & Londono, J. L. (1997). Asset inequality matters: An assessment of the world bank's approach to poverty reduction. *The American Economic Review, 87*(2), 32-37.

Blankenau, W., Cassou, S. P., & Ingram, B. (2007). Allocating government education expenditures across k-12 and college education. *Economic Theory, 31*(1), 85-112. doi:10.1007/s00199-006-0084-8

Boldrin, M. (1995). Labor contracts and business cycles. *Journal of Political Economy, 103*(5), 972-1004. doi:10.1086/262010

Boldrin, M. (2005). Public education and capital accumulation. *Research in Economics, 59*(2), 85-109.

Bollard, A., Klenow, P. J., & Sharma, G. (2013). India's mysterious manufacturing miracle. *Review of Economic Dynamics, 16*(1), 59-85.

Boulhol, H. (2004). *Technology differences, institutions and economic growth: a conditional conditional convergence* (Vol. 2): CEPII.

Bräuninger, M., & Vidal, J. P. (2000). Private versus public financing of education and endogenous growth. *Journal of Population Economics, 13*(3), 387-401. doi:10.1007/s001480050143

Brianzoni, S., Mammana, C., & Michetti, E. (2012). Variable elasticity of substituition in a discrete time Solow-Swan growth model with differential saving. *Chaos, Solitons and Fractals, 45*(1), 98.

Castelló-Climent, A., & Doménech, R. (2008). Human capital inequality, life expectancy and economic growth. *The Economic Journal, 118*(528), 653-677. doi:10.2307/20108816

Catrinescu, N., Leon-Ledesma, M., Piracha, M., & Quillin, B. (2009). Remittances, institutions, and economic growth. *World Development*, *37*(1), 81-92.

Chakraborty, S. (2004). Endogenous lifetime and economic growth. *Journal of Economic Theory, 116*(1), 119-137.

Chakraborty, S., & Das, M. (2005). Mortality, human capital and persistent inequality. *Journal of Economic Growth, 10*(2), 159-192. doi:10.1007/s10887-005-1670-5

Chami, R., Barajas, A., Cosimano, T., Fullenkamp, C., Gapen, M., & Montiel, P. (2008). *Macroeconomic consequences of remittances*: International Monetary Fund Washington, DC.

Chang, H. J. (2011). Institutions and economic development: theory, policy and history. *Journal of Institutional Economics, 7*(4), 473-498. doi:10.1017/S1744137410000378

Chang, H., Cheng, M. A., & Das, S. (2004). Hospital ownership and operating efficiency: Evidence from Taiwan. *European Journal of Operational Research, 159*(2), 513-527. doi:10.1016/S0377-2217(03)00412-0

Chenery, H. B. (1959). The interdependence of investment decisions. *The Allocation of Economic Resources*, 82-120.

Cheng, T., & Vahid, F. (2010). *Demand for hospital care and private health insurance in a mixed public-private system: empirical evidence using a simultaneous equation modeling approach* (No. 22/11). Melbourne Institute Working Paper Series, No. 22/11.

Clarke, G. R. G. (1995). More evidence on income distribution and growth. *Journal of Development Economics, 47*(2), 403-427. doi:10.1016/0304-3878(94)00069-O

Coleman, J. S., Hoffer, T., & Kilgore, S. (1982). *High school achievement: public, Catholic, and private schools compared*. New York: Basic Books.

Crafts, N. (2004). Globalisation and economic growth: a historical perspective. *The World Economy, 27*(1), 45-58.

Dahlman, C. J., Ross-Larson, B., & Westphal, L. E. (1987). Managing technological development: Lessons from the newly industrializing countries. *World Development, 15*(6), 759-775.

Das, M. (2007). Persistent inequality: An explanation based on limited parental altruism. *Journal of Development Economics, 84*(1), 251-270. doi:10.1016/j.jdeveco.2005.10.006

De Gregorio, J., & Guidotti, P. E. (1995). Financial development and economic growth. *World Development, 23*(3), 433-448. doi:10.1016/0305-750X(94)00132-I

de la Croix, D., & Doepke, M. (2004). Public versus private education when differential fertility matters. *Journal of Development Economics, 73*(2), 607-629. doi:10.1016/j.jdeveco.2003.05.005

de la Croix, D., & Michel, P. (2002). *A theory of economic growth : Dynamics and policy in overlapping generations*. Cambridge: Cambridge University Press.

De la Fuente, A. (1997). The empirics of growth and convergence: a selective review. *Journal of Economic Dynamics and Control, 21*(1), 23-73.

de La Grandville, O. (1989). In quest of the Slutsky diamond. *The American Economic Review, 79*(3), 468-481.

Deaton, A. (2003). Health, inequality, and economic development. *Journal of Economic Literature, 41*(1), 113-158. doi:10.1257/002205103321544710

Deininger, K., & Squire, L. (1998). New ways of looking at old issues: inequality and growth. *Journal of Development Economics, 57*(2), 259-287. doi:10.1016/S0304-3878(98)00099-6

Desdoigts, A. (1999). Patterns of economic development and the formation of clubs. *Journal of Economic Growth, 4*(3), 305-330. doi:10.1023/A:1009858603947

Di Gropello, E. (2012). *Putting higher education to work: skills and research for growth in East Asia* (0821384902, 9780821384909).

Di Maria, C., & Lazarova, E. A. (2012). Migration, human capital formation, and growth: an empirical investigation. *World Development, 40*(5), 938-955. doi:10.1016/j.worlddev.2011.11.011

Di Maria, C., & Stryszowski, P. (2009). Migration, human capital accumulation and economic development. *Journal of Development Economics, 90*(2), 306-313. doi:http://dx.doi.org/10.1016/j.jdeveco.2008.06.008

Diamond, J. (1997). Guns, germs and steel: The fates of human societies. *New York: W. W. Norton*.

Diamond, P. A. (1965). National debt in a neoclassical growth model. *The American Economic Review, 55*(5), 1126-1150. doi:10.2307/1809231

Docquier, F., Lohest, O., & Marfouk, A. (2007). Brain drain in developing countries. *The World Bank Economic Review, 21*(2), 193-218. doi:10.1093/wber/lhm008

Dollar, D., & Kraay, A. (2003). Institutions, trade, and growth. *Journal of Monetary Economics, 50*(1), 133-162.

Duffy, J., & Papageorgiou, C. (2000). A cross-country empirical investigation of the aggregate production function specification. *Journal of Economic Growth, 5*(1), 87-120. doi:10.1023/A:1009830421147

Durlauf, S. N. (1995). Multiple regimes and cross-country growth behaviour. *Journal of Applied Econometrics, 10*(4), 365-384. doi:10.1002/jae.3950100404

Durlauf, S. N. (1996). A theory of persistent income inequality. *Journal of Economic Growth, 1*(1), 75-93.

Easterly, W., & Levine, R. (2003). Tropics, germs, and crops: how endowments influence economic development. *Journal of Monetary Economics, 50*(1), 3-39.

Epple, D., & Romano, R. E. (1996a). Ends against the middle: determining public service provision when there are private alternatives. *Journal of Public Economics, 62*(3), 297-325.

Epple, D., & Romano, R. E. (1996b). Public provision of private goods. *The Journal of Political Economy, 104*(1), 57-84. doi:10.1086/262017

European Commission. (2014). Government expenditure by function – COFOG. Retrieved from European Commission- Eurostat website: http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Government_expenditure_by_function_%E2%80%93_COFOG#>

Fabbri, D., & Monfardini, C. (2009). Rationing the public provision of healthcare in the presence of private supplements: evidence from the Italian NHS. *Journal of Health Economics, 28*(2), 290-304.

Fagerberg, J. (1994). Technology and international differences in growth rates. *Journal of Economic Literature, 32*(3), 1147-1175.

Färe, R., & Yoon, B. J. (1981). Variable elasticity of substitution in urban housing production. *Journal of Urban Economics, 10*(3), 369-374. doi:10.1016/0094-1190(81)90007-3

Fernández-Ballesteros, R., Zamarrón, M. D., Díez-Nicolás, J., López-Bravo, M. D., Molina, M. Á., & Schettini, R. (2011). Productivity in old age. *Research on Aging, 33*(2), 205-226.

Fields, G. S., & Yoo, G. (2000). Falling labor income inequality in Korea's economic growth: patterns and underlying causes. *Review of Income and Wealth, 46*(2), 139-159. doi:10.1111/j.1475-4991.2000.tb00952.x

Filmer, D., Hammer, J. S., & Pritchett, L. H. (2000). Weak links in the chain: A diagnosis of health policy in poor countries. *The World Bank Research Observer, 15*(2), 199-224.

Gallup, J. L., Sachs, J. D., & Mellinger, A. D. (1999). Geography and economic development. *International Regional Science Review, 22*(2), 179-232.

Galor, O. (1996). Convergence? Inferences from theoretical models. *The Economic Journal, 106*(437), 1056-1069.

Galor, O. (2007). *Discrete Dynamical Systems*: Springer-Verlag.

Galor, O. (2011). *Inequality, human capital formation and the process of development* (No. w17058). National Bureau of Economic Research.

Galor, O., & Moav, O. (2004). From physical to human capital accumulation: Inequality and the process of development. *Review of Economic Studies, 71*(249), 1001-1026. doi:10.1111/j.1467-937X.2004.00312.x

Galor, O., & Tsiddon, D. (1997). The distribution of human capital and economic growth. *Journal of Economic Growth, 2*(1), 93-124.

Galor, O., & Zeira, J. (1993). Income distribution and macroeconomics. *The Review of Economic Studies, 60*(1), 35-52.

Ghosh, M., Ghoshray, A., & Malki, I. (2013). Regional divergence and club convergence in india. *Economic Modelling, 30*, 733-742. doi:10.1016/j.econmod.2012.10.008

Glaeser, E. L., La Porta, R., Lopez-De-Silanes, F., & Shleifer, A. (2004). Do institutions cause growth? *Journal of Economic Growth, 9*(3), 271-303. doi:10.1023/B:JOEG.0000038933.16398.ed

Glomm, G., & Ravikumar, B. (1998). Increasing returns, human capital, and the Kuznets curve. *Journal of Development Economics, 55*(2), 353-367.

Glomm, G., & Ravikumar, B. (1997). Productive government expenditures and long-run growth. *Journal of Economic Dynamics and Control, 21*(1), 183-204.

Glomm, G. (1997). Parental choice of human capital investment. *Journal of Development Economics, 53*(1), 99-114. doi:10.1016/S0304-3878(97)00010-2

Glomm, G., & Ravikumar, B. (1992). Public versus private investment in human capital: Endogenous growth and income inequality. *The Journal of Political Economy, 100*(4), 818-834. doi:10.1086/261841

Goldhaber, D. (1999). An endogenous model of public school expenditures and private school enrollment. *Journal of Urban Economics, 46*(1), 106-128.

Goldin, C. D., & Katz, L. F. (2009). *The race between education and technology*: Harvard University Press.

Gollin, D. (2002). Getting income shares right. *Journal of Political Economy, 110*(2), 458-474. doi:10.1086/338747

Gottschalk, P., & Joyce, M. (1998). Cross-national differences in the rise in earnings inequality: Market and institutional factors. *The Review of Economics and Statistics, 80*(4), 489-502. doi:10.1162/003465398557672

Gottschalk, P., & Smeeding, T. M. (1997). Cross-national comparisons of earnings and income inequality. *Journal of Economic Literature*, *35*(2), 633-687. Gouveia, M. (1997). Majority rule and the public provision of a private good. *Public Choice, 93*(3), 221-244. doi:10.1023/A:1017929005280

Grossman, M. (1972). The demand for health: a theoretical and empirical investigation. *NBER Books*.

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth, 17*(4), 267-321. doi:10.1007/s10887-012-9081-x

Hardie, E. A., & Critchley, C. R. (2008). Public perceptions of Australia's doctors, hospitals and health care systems. *The Medical Journal of Australia, 189*(4), 210.

Hazan, M., & Zoabi, H. (2006). Does longevity cause growth? A theoretical critique. *Journal of Economic Growth, 11*(4), 363-376.

Hulten, C. R. (2001). Total factor productivity: a short biography *New developments in productivity analysis* (pp. 1-54): University of Chicago Press

IHME. (2014). Adult mortality rates by country and sex mapped over time (Global), 1970-2010.

ILO. (2013). *Global Wage Report 2012-13: Wages and Equitable Growth*: International Labour Office.

ILO. (2014). *Global Wage Report: 2014/2015*: International Labor Office.

ILOSTAT. (2015). *Global Wage Report Collection*.

Irmen, A. (2003). On the growth effects of the elasticity of substitution. *mimeo, University of Mannheim.*

Irmen, A. (2011). Steady-state growth and the elasticity of substitution. *Journal of Economic Dynamics and Control, 35*(8), 1215-1228. doi:http://dx.doi.org/10.1016/j.jedc.2011.04.002

Irmen, A., & Klump, R. (2009). Factor substitution, income distribution and growth in a generalized neoclassical model. *German Economic Review, 10*(4), 464-479. doi:10.1111/j.1468-0475.2009.00491.x

Irmen, A., & Maußner, A. (2014). Essential inputs and unbounded output: An alternative characterization of the neoclassical production function. *CESIFO working paper, No. 5126.*

Islam, N. (1995). Growth empirics: A panel data approach. *The Quarterly Journal of Economics, 110*(4), 1127-1170. doi:10.2307/2946651

James, E. (1993). Why do different countries choose a different public-private mix of educational services? *The Journal of Human Resources*, *28*(3), 571-592.

James, T. (2004). The right to education. *Economic Affairs, 24*(3), 75-75. doi:10.1111/j.1468-0270.2004.t01-1-00498.x

Jerzmanowski, M. (2006). Empirics of hills, plateaus, mountains and plains: A Markov-switching approach to growth. *Journal of Development Economics, 81*(2), 357-385. doi:http://dx.doi.org/10.1016/j.jdeveco.2005.05.005

Jimenez, E., & Lockheed, M. E. (1995). *Public and private secondary education in developing countries: A comparative study* (Vol. 309). World Bank Publications.

Johannesson, M., Johansson, P. O., Kriström, B., Borgquist, L., & Jönsson, B. (1993). Willingness to pay for lipid lowering: a health production function approach. *Applied Economics, 25*(8), 1023.

Jones, C. I. (1997). On the Evolution of the World Income Distribution. *The Journal of Economic Perspectives, 11*(3), 19-36. doi:10.1257/jep.11.3.19

Jones, C. I. (2001). Was an industrial revolution inevitable? Economic growth over the very long run. *Advances in Macroeconomics, 1*(2).

Jones, C. I. (2005). The shape of production functions and the direction of technical change. *The Quarterly Journal of Economics, 120*(2), 517-549. doi:10.1093/qje/120.2.517

Jones, R. W. (2008). The aggregate elasticity of factor substitution with middle products. *Journal of Macroeconomics, 30*(2), 633-640.

Kadiyala, K. R. (1972). Production functions and elasticity of substitution. *Southern Economic Journal, 38*(3), 281-284.

Kaganovich, M., & Zilcha, I. (1999). Education, social security, and growth. *Journal of Public Economics, 71*(2), 289-309. doi:http://dx.doi.org/10.1016/S0047-2727(98)00073-5

Kahn, H. (1979). *World economic development: 1979 and beyond*. Boulder, Colorado: Westview Press.

Kaplinsky, R., & Messner, D. (2008). Introduction: the impact of Asian drivers on the developing world. *World Development, 36*(2), 197-209.

Karabarbounis, L., & Neiman, B. (2014). The global decline of the labor share. *The Quarterly Journal of Economics, 129*(1), 61-108.

Karagiannis, G., Palivos, T., & Papageorgiou, C. (2005). Variable elasticity of substitution and economic growth: Theory and evidence (pp. 21-37). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kazi, U. A. (1980). The variable elasticity of substitution production function: A case study for Indian manufacturing industries. *Oxford Economic Papers, 32*(1), 163-175.

Kenny, C., & Williams, D. (2001). What do we know about economic growth? Or, why don't we know very much? *World Development, 29*(1), 1-22. doi:10.1016/S0305-750X(00)00088-7

Kerekes, M. (2012). Growth miracles and failures in a Markov switching classification model of growth. *Journal of Development Economics, 98*(2), 167-177.

Kim, S., & Lee, J. H. (2010). Private tutoring and demand for education in south korea. *Economic Development and Cultural Change, 58*(2), 259-296. doi:10.1086/648186

Kimenyi, M. S. (2011). Institutions and development: the primacy of microanalysis. *Journal of Institutional Economics, 7*(4), 549-553. doi:10.1017/S1744137411000105

Kitagawa, A., & Shibata, A. (2005). Endogenous growth cycles in an overlapping generations model with investment gestation lags. *Economic Theory, 25*(3), 751-762.

Klump, R., & De La Grandville, O. (2000). Economic growth and the elasticity of substitution: Two theorems and some suggestions. *The American Economic Review, 90*(1), 282-291. doi:10.1257/aer.90.1.282

Klump, R., & Preissler, H. (2000). CES production functions and economic growth. *The Scandinavian Journal of Economics, 102*(1), 41-56. doi:10.1111/1467-9442.00183

KPMG. (2012). *The state of health care in Africa*. Retrieved from http://www.kpmg.com/Africa/en/IssuesAndInsights/Articles-Publications/Documents/The-State-of-Healthcare-in-Africa.pdf

Krusell, P., & Ríos-Rull, J. V. (1996). Vested interests in a positive theory of stagnation and growth. *The Review of Economic Studies, 63*(2), 301-329.

Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*, 1-28.

Lahiri, R. (2002). The inflation tax, variable time preference, and the business cycle. *Macroeconomic Dynamics, 6*(04), 496-522.

Lahiri, R., & Richardson, E. (2009). *Health, endogenous time preference, and inequality: a political economy perspective.* Paper presented at the Proceedings of Econometric Society Australasian Meeting in 2009.

Le, T. (2009). Trade, remittances, institutions, and economic growth. *International Economic Journal, 23*(3), 391-408.

Lee, C. J., Lee, H., & Jang, H. M. (2010). The history of policy responses to shadow education in South Korea: Implications for the next cycle of policy responses. *Asia Pacific Education Review, 11*(1), 97-108.

León-Ledesma, M. A., McAdam, P., & Willman, A. (2010). Identifying the elasticity of substitution with biased technical change. *The American Economic Review, 100*(4), 1330-1357. doi:10.1257/aer.100.4.1330

Leung, M., & Wang, Y. (2010). Endogenous health care, life expectancy and economic growth. *Pacific Economic Review, 15*(1), 11-31.

Levin, H. M., & Belfield, C. R. (2003). The marketplace in education. *Review of Research in Education, 27*, 183-219.

Levine, R., & King, R. G. (1993). Finance, entrepreneurship and growth: Theory and evidence. *Journal of Monetary Economics, 32*(3), 513-542.

Levine, R., & Zervos, S. (1998). Stock markets, banks, and economic growth. *The American Economic Review, 88*(3), 537-558.

Levy, V. (1981). Total factor productivity, non-neutral technical change and economic growth: A parametric study of a developing economy. *Journal of Development Economics, 8*(1), 93-109. doi:http://dx.doi.org/10.1016/0304-3878(81)90047-X

Lewis, W. A. (1954). Economic Development with Unlimited Supplies of Labour *Manchester School of Economic and Social Studies, 22*(2), 139.

Li, S. M., Moslehi, S., & Yew, S. L. (2012). *Public-private mix of health expenditure: A political economy approach and a quantitative exercise* (No. 11). Monash University Discussion Papers.

Lien, D., & Wang, Y. (2005). Brain drain or brain gain: A revisit. *Journal of Population Economics, 18*(1), 153-163. doi:10.1007/s00148-003-0174-x

Limão, N., & Venables, A. J. (2001). Infrastructure, geographical disadvantage, transport costs, and trade. *The World Bank Economic Review, 15*(3), 451-479. doi:10.2307/3990110

Lipsey, R. G., & Carlaw, K. (2000). What does total factor productivity measure? *International Productivity Monitor, 1*, 31-40.

Lopez, R., Thomas, V., & Wang, Y. (1998). *Addressing the education puzzle: the distribution of education and economic reform* (No. 2031). The World Bank.

Lu, Y. C., & Fletcher, L. B. (1968). A generalization of the CES production function. *The Review of Economics and Statistics, 50*(4), 449-452.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics, 22*(1), 3-42.

Lülfesmann, C., & Myers, G. M. (2010). Two-tier public provision: Comparing public systems. *Journal of Public Economics, 95*(11-12), 1263. doi:10.1016/j.jpubeco.2010.08.002

MacKellar, L., Wörgötter, A., & Wörz, J. (2000). *Economic development problems of landlocked countries* (No. 14). Institute for Advanced Studies.

Macpherson, D. A. (1990). Trade unions and labor's share in U.S. manufacturing industries. *International Journal of Industrial Organization, 8*(1), 143-151. doi:http://dx.doi.org/10.1016/0167-7187(89)90038-6

Mallick, D. (2010). Capital–labor substitution and balanced growth. *Journal of Macroeconomics, 32*(4), 1131-1142. doi:10.1016/j.jmacro.2010.04.003

Mallick, D. (2012a). The role of capital-labour substitution in economic growth. *Indian Growth and Development Review, 5*(1), 89-101. doi:10.1108/17538251211224150

Mallick, D. (2012b). The role of the elasticity of substitution in economic growth: a cross-country investigation. *Labour Economics, 19*(5), 682-694. doi:10.1016/j.labeco.2012.04.003

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A Contribution to the empirics of economic growth. *The Quarterly Journal of Economics, 107*(2), 407-437. doi:10.2307/2118477

Masiye, F. (2007). Investigating health system performance: an application of data envelopment analysis to Zambian hospitals. *BMC Health Services Research, 7*(1), 58-58. doi:10.1186/1472-6963-7-58

Mata, T., & Louçã, F. (2009). The Solow residual as a black box: attempts at integrating business cycle and growth theories. *History of Political Economy, 41*(Suppl 1), 334-355.

McDonald, S., & Zhang, J. (2012). Income inequality and economic growth with altruistic bequests and human capital investment. *Macroeconomic Dynamics, 16*(S3), 1-24. doi:10.1017/S136510051000101X

McIntosh, J. (1986). Economic growth and technical change in Britain 1950–1978. *European Economic Review, 30*(1), 117-128. doi:http://dx.doi.org/10.1016/0014-2921(86)90035-8

Michel, P., & Venditti, A. (1997). Optimal growth and cycles in overlapping generations models. *Economic Theory, 9*(3), 511-528.

Milanovic, B., & Yitzhaki, S. (2002). Decomposing world income distribution: Does the world have a middle class? *Review of Income and Wealth, 48*(2), 155-178. doi:10.1111/1475-4991.00046

Mincer, J. (1958). Investment in human capital and personal income distribution. *The Journal of Political Economy, 66*(4), 281-302.

Mirowsky, J., & Ross, C. E. (1998). Education, personal control, lifestyle and health: A human capital hypothesis. *Research on Aging, 20*(4), 415-449. doi:10.1177/0164027598204003

Miyagiwa, K., & Papageorgiou, C. (2003). Elasticity of substitution and growth: Normalized CES in the Diamond model. *Economic Theory, 21*(1), 155-165. doi:10.1007/s00199-002-0268-9

Miyagiwa, K., & Papageorgiou, C. (2007). Endogenous aggregate elasticity of substitution. *Journal of Economic Dynamics and Control, 31*(9), 2899-2919. doi:10.1016/j.jedc.2006.06.009

Mountford, A. (1997). Can a brain drain be good for growth in the source economy? *Journal of Development Economics, 53*(2), 287-303. doi:10.1016/S0304-3878(97)00021-7

Nelson, R. R., & Phelps, E. S. (1966). Investment in humans, technological diffusion, and economic growth. *The American Economic Review*, 69-75.

Ng, Y. K. (2003). *Welfare economics: towards a more complete analysis*. New York: Palgrave Macmillan.

North, D. C. (1989). Institutions and economic growth: An historical introduction. *World Development, 17*(9), 1319-1332. doi:10.1016/0305-750X(89)90075-2

Odedokun, M. O. (1996). Alternative econometric approaches for analysing the role of the financial sector in economic growth: Time-series evidence from LDCs. *Journal of Development Economics, 50*(1), 119-146. doi:10.1016/0304-3878(96)00006-5

OECD. (2012a). *Is Increasing Private Expenditure, Especially in Tertiary Education, Associated with Less Public Funding and Less Equitable Access?* : OECD Publishing.

OECD. (2012b). *Labour Losing to Capital: What Explains the Declining Labour Share?* : OECD Publishing.

OECD. (2013). *OECD Economic Surveys: France 2013*. Paris: OECD Publishing.

Olsson, O., & Hibbs, D. A. (2005). Biogeography and long-run economic development. *European Economic Review, 49*(4), 909-938. doi:10.1016/j.euroecorev.2003.08.010

Osang, T., & Sarkar, J. (2008). Endogenous mortality, human capital and economic growth. *Journal of Macroeconomics, 30*(4), 1423-1445. doi:10.1016/j.jmacro.2007.09.002

Pack, H. (1994). Endogenous growth theory: Intellectual appeal and empirical shortcomings. *The Journal of Economic Perspectives, 8*(1), 55-72. doi:10.1257/jep.8.1.55

Palivos, T., & Karagiannis, G. (2010). The elasticity of substitution as an engine of growth. *Macroeconomic Dynamics, 14*(5), 617-628. doi:10.1017/S1365100509000479

Persson, T., & Tabellini, G. (1994). Is inequality harmful for growth? *The American Economic Review, 84*(3), 600-621.

Piketty, T., & Goldhammer, A. (2014). *Capital in the twenty-first century*. Cambridge Massachusetts: The Belknap Press of Harvard University Press.

Pitchford, J. (1960). Growth and the elasticity of factor substitution. *Economic Record, 36*(76), 491-504.

Prescott, E. C. (1986). Theory ahead of business-cycle measurement. *Carnegie-Rochester Conference Series on Public Policy, 25*(1), 11-44. doi:10.1016/0167-2231(86)90035-7

Pritchett, L. H. (1997). Divergence, big time. *The Journal of Economic Perspectives, 11*(3), 3-17. doi:10.1257/jep.11.3.3

Pritchett, L. H. (2000). Understanding patterns of economic growth: Searching for hills among plateaus, mountains, and plains. *The World Bank Economic Review, 14*(2), 221-250.

Przeworski, A. (2004). The last instance: Are institutions the primary cause of economic development? *European Journal of Sociology / Archives Européennes de Sociologie, 45*(2), 165-188. doi:10.1017/S0003975604001419

Przeworski, A., & Limongi, F. (1993). Political regimes and economic growth. *The Journal of Economic Perspectives (1986-1998), 7*(3), 51.

Quah, D. T. (1996). Empirics for economic growth and convergence. *European Economic Review, 40*(6), 1353-1375. doi:10.1016/0014-2921(95)00051-8

Quah, D. T. (1997). Empirics for growth and distribution: stratification, polarization, and convergence clubs. *Journal of Economic Growth, 2*(1), 27-59. doi:10.1023/A:1009781613339

Rao, B. B., & Hassan, G. M. (2011). A panel data analysis of the growth effects of remittances. *Economic Modelling, 28*(1), 701-709.

Ravallion, M. (1997). Can high-inequality developing countries escape absolute poverty? *Economics Letters, 56*(1), 51-57.

Ravallion, M. (2001). Growth, inequality and poverty: looking beyond averages. *World Development, 29*(11), 1803-1815.

Ray, D., & Streufert, P. A. (1993). Dynamic equilibria with unemployment due to undernourishment. *Economic Theory, 3*(1), 61-85. doi:10.1007/BF01213692

Revankar, N. S. (1971). A class of variable elasticity of substitution production functions. *Econometrica, 39*(1), 61-71.

Rodriguez, C. A. (1975b). On the welfare aspects of international migration. *Journal of Political Economy, 83*(5), 1065-1072. doi:10.1086/260378

Rodriguez, F., & Jayadev, A. (2010). The declining labor share of income. *Journal of Globalization and Development, 3*(2), 1-18.

Rodrik, D., & Chen, T. J. (1998). *TFPG controversies, institutions and economic performance in East Asia* (pp. 79-105). Palgrave Macmillan UK.

Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions rule: The primacy of institutions over geography and integration in economic development. *Journal of Economic Growth, 9*(2), 131-165. doi:10.1023/B:JOEG.0000031425.72248.85

Romer, P. M. (1986). Increasing returns and long-run growth. *The Journal of Political Economy, 94*(5), 1002-1037. doi:10.1086/261420

Romer, P. M. (1987). Growth based on increasing returns due to specialization. *The American Economic Review,* 56-62.

Romer, P. M. (1994). The origins of endogenous growth. *The Journal of Economic Perspectives, 8*(1), 3-22. doi:10.1257/jep.8.1.3

Roubini, N., & Sala-i-Martin, X. (1992). Financial repression and economic growth. *Journal of Development Economics*, *39*(1), 5-30. doi:10.1016/0304-3878(92)90055-E

Sachs, J. D. (2001). *Tropical underdevelopment* (No. w8119). National Bureau of Economic Research.

Sala-i-Martin, X. (2002). *15 years of new growth economics: what have we learnt?* (No. 172). Central Bank of Chile Working Papers

Sala-i-Martin, X. (2006). The world distribution of income: falling poverty and... convergence, period. *The Quarterly Journal of Economics*, 351-397.

Sander, W. (1999). Private schools and public school achievement. *Journal of Human Resources, 34*(4), 697-709.

Sarkar, J.. (2008). Mortality, fertility, and persistent income inequality. *Southern Economic Journal,75*(2), 332–350.

Sato, R., & Hoffman, R. F. (1968). Production functions with variable elasticity of factor substitution: Some analysis and testing. *The Review of Economics and Statistics*, *50*(4), 453-460.

Schultz, T. W. (1961). Investment in human capital. *The American Economic Review,* 1-17.

Schumpeter, J. A. (1934). *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle* (Vol. 55): Transaction publishers.

Shin, I. (2012). Income inequality and economic growth. *Economic Modelling, 29*(5), 2049-2057.

Sirmans, C. F., Kau, J. B., & Lee, C. F. (1979). The elasticity of substitution in urban housing production: A VES approach. *Journal of Urban Economics, 6*(4), 407-415. doi:10.1016/0094-1190(79)90021-4

Skeldon, R. (2008). International migration as a tool in development policy: A passing phase? *Population and Development Review, 34*(1), 1-18. doi:10.1111/j.1728-4457.2008.00203.x

Skirbekk, V. (2004). Age and individual productivity: A literature survey. *Vienna yearbook of population research*, 133-153.

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, *70*(1), 65-94.

Solow, R. M. (1997). It ain't the things you don't know that hurt you, it's the things you know that ain't so. *The American Economic Review, 87*(2), 107-108.

Soroka, S. N. (2007). *Canadian perceptions of the health care system*. Health Council of Canada.

Stark, O. (2004). Rethinking the brain drain. *World Development, 32*(1), 15-22. doi:http://dx.doi.org/10.1016/j.worlddev.2003.06.013

Stark, O., Helmenstein, C., & Prskawetz, A. (1997). A brain gain with a brain drain. *Economics Letters, 55*(2), 227-234. doi:10.1016/S0165-1765(97)00085-2

Stark, O., & Wang, Y. (2002). Inducing human capital formation: migration as a substitute for subsidies. *Journal of Public Economics, 86*(1), 29-46. doi:http://dx.doi.org/10.1016/S0047-2727(01)00104-9

Stevens, C. M. (1977). Health and economic development: Longer-run view. *Social Science & Medicine (1967), 11*(17–18), 809-817. doi:http://dx.doi.org/10.1016/0037-7856(77)90120-2

Strawczynski, M. (1993). Income uncertainty, bequests and annuities. *Economics Letters, 42*(2), 155-158.

Tallman, E. W., & Wang, P. (1994). Human capital and endogenous growth evidence from Taiwan. *Journal of Monetary Economics, 34*(1), 101-124. doi:http://dx.doi.org/10.1016/0304-3932(94)01152-4

Tamura, R. F. (1991). Income convergence in an endogenous growth model. *The Journal of Political Economy, 99*(3), 522-540.

Tamura, R. F. (2006). Human capital and economic development. *Journal of Development Economics, 79*(1), 26-72.

Temple, J. (1999). The new growth evidence. *Journal of Economic Literature, 37*(1), 112-156. doi:10.1257/jel.37.1.112

Tuohy, C. H., Flood, C. M., & Stabile, M. (2004). How does private finance affect public health care systems? Marshaling the evidence from OECD nations. *Journal of Health Politics, Policy and Law, 29*(3), 359-396. doi:10.1215/03616878-29-3-359

Varvarigos, D., & Zakaria, I. Z. (2013). Endogenous fertility in a growth model with public and private health expenditures. *Journal of Population Economics, 26*(1), 67-85. doi:10.1007/s00148-012-0412-1

Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin, 122*(3), 231-249. doi:10.1037/0033-2909.122.3.231

Weber, M. (1930). *The Protestant ethic and the spirit of capitalism*. London: Unwin University Books.

Whitlock, G. (2012). *Mortality trends.* Retrieved from http://www.mortality-trends.org/info.html

Wilson, K., & Rosenberg, M. W. (2004). Accessibility and the Canadian health care system: Squaring perceptions and realities. *Health Policy, 67*(2), 137-148. doi:10.1016/S0168-8510(03)00101-5

Wongsurawat, W. (2011). Education reform and the academic performance of public and private secondary school students in Thailand. *Educational Research for Policy and Practice, 10*(1), 17-28. doi:10.1007/s10671-010-9093-2

World Bank. (2011). *Technical Note: Inequality measures and their decompositions.* Retrieved from http://siteresources.worldbank.org/INTPA/Resources/tn_measuring_inequality.pdf

World Bank. (2014). *Private enrollment as a percentage of total secondary school enrolment*. Retrieved from: http://data.worldbank.org/indicator/SE.SEC.PRIV.ZSWorld Bank. (2015). *GDP growth (annual %)*. Retrieved from: http://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG

World Bank. (2016). *World DataBank*. Retrieved from: http://databank.worldbank.org/data/home.aspx

Yaari, M. E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies, 32*(2), 137-150.

Yanikkaya, H. (2003). Trade openness and economic growth: a cross-country empirical investigation. *Journal of Development Economics, 72*(1), 57-89. doi:10.1016/S0304-3878(03)00068-3

Yip, C. K., & Xue, J. (2013). Aggregate elasticity of substitution and economic growth: A synthesis. *Journal of Macroeconomics, 38*(1), 60-75. doi:10.1016/j.jmacro.2013.07.001

Young, A. T. (2010). One of the things we know that ain't so: is US labor's share relatively stable? *Journal of Macroeconomics, 32*(1), 90-102. doi:10.1016/j.jmacro.2009.09.008

Yuhn, K. H. (1991). Economic growth, technical change biases, and the elasticity of substitution: A test of the de la Grandville hypothesis. *The Review of Economics and Statistics, 73*(2), 340-346.

Zahid, B., Muhammad Iqbal, M., & Tariq, M. (2011). Comparison of quality indicators in public and private secondary schools of Hazara region, Pakistan. *Interdisciplinary Journal of Contemporary Research In Business, 2*(11), 157.

Zhou, Y. (2008). Synchronizing export orientation with import substitution: Creating competitive indigenous high-tech companies in China. *World Development, 36*(11), 2353-2370. doi:10.1016/j.worlddev.2007.11.013

Ziesemer, T. H. (2012). Worker remittances, migration, accumulation and growth in poor developing countries: Survey and analysis of direct and indirect effects. *Economic Modelling, 29*(2), 103-118.

Zweifel, P., Breyer, F., & Kifmann, M. (2009). *Health economics*. New York: Springer Berlin Heidelberg.

Appendices

Appendix A

Proof of Lemma 1 in Chapter 3

From (3.2.3), the derivative of MP_K with respect to b is given by:

$$\frac{\partial MP_K}{\partial b} = \frac{ak^a(1-a)(1+a+abk)}{(1+abk)^{a+1}} \quad \text{(A1)}$$

From (A1) above, we can see that the interest rate is increasing in b when $b > 0$. When $b < 0$, the interest rate is increasing in b if $k < \frac{(1+a)}{a|b|}$. This condition is always satisfied in the range $k < 1/|b|$ which was established earlier.

From (3.2.5), it is clear that a higher elasticity of substitution, captured by a higher value of the parameter b yields a lower MP_L and thereby a lower wage rate.

Appendix B

Proof of Proposition 1 in Chapter 3

First, we derive the condition for stability of a given steady state. We linearise equation (3.3.11) around the steady state. The linearised system can be approximated by:

$$k_{t+1} = f\left(\bar{k}\right) + f'\left(\bar{k}\right)\left(k_t - \bar{k}\right) \tag{B1}$$

From (3.12), we have:

$$f'\left(\bar{k}\right) = \frac{\beta a(1-a)\bar{k}^{a-1}}{(1+\beta)\left(1+ab\bar{k}\right)^{(1+a)}} \tag{B2}$$

Therefore, the linearised system can then be written as:

$$k_{t+1} = \frac{\beta a(1-a)\bar{k}^{a-1}}{(1+\beta)\left(1+ab\bar{k}\right)^{(1+a)}}\left[\left(1+ab\bar{k}\right)\bar{k} - a\right] + \frac{\beta a(1-a)\bar{k}^{a-1}}{(1+\beta)\left(1+ab\bar{k}\right)^{(1+a)}} k_t \tag{B3}$$

For the steady state capital stock to be locally stable, the following condition must be satisfied:

$$\frac{\beta a(1-a)\bar{k}^{a-1}}{(1+\beta)\left(1+ab\bar{k}\right)^{(1+a)}} < 1 \tag{B4}$$

By rewriting (3.3.12) as:

$$\bar{k}^{1-a}\left(1+ab\bar{k}\right)^{a} = \frac{\beta(1-a)}{(1+\beta)} \tag{B5}$$

The local stability condition simplifies to:

$$1 + ab\bar{k} > a \tag{B6}$$

When $b > 0$, (B6) simplifies to:

$$\bar{k} > \frac{-(1-a)}{ab} \tag{B7}$$

The inequality given by (B7) is always satisfied because $\bar{k} \geq 0$. So, when $b > 0$, the resulting steady state is always locally stable.

When $b < 0$, the local stability condition simplifies to:

$$\bar{k} < \frac{(1-a)}{a|b|} \tag{B8}$$

Before looking at when this local stability condition is satisfied, the number of steady states associated with different parameter values must be examined. From (B5), let $M = \bar{k}^{1-a}\left(1 + ab\bar{k}\right)^{a} = \dfrac{\beta(1-a)}{(1+\beta)}$

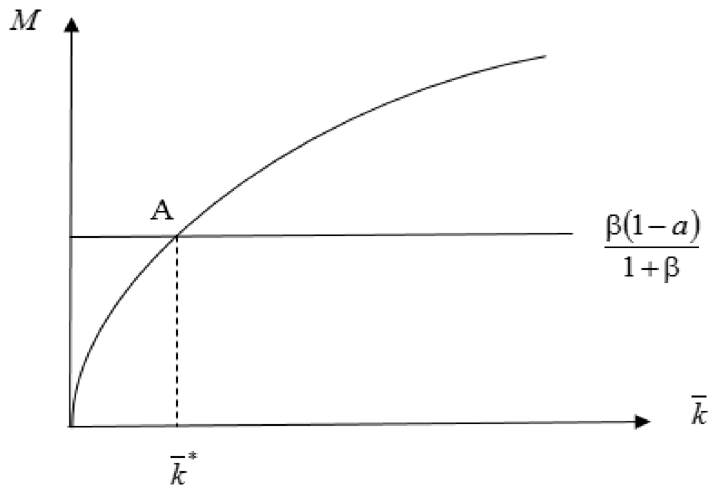
The first and the second derivatives of M are given by:

$$\frac{dM}{d\bar{k}} = \frac{1 - a + ab\bar{k}}{\bar{k}^{a}\left(1 + ab\bar{k}\right)^{1-a}} \tag{B9}$$

$$\frac{d^2M}{d\bar{k}^2} = \frac{-a(1-a)}{\bar{k}^{a+1}\left(1 + ab\bar{k}\right)^{3a+2}} \tag{B10}$$

When $b > 0$, (B9) is always positive, while for $b < 0$, the function M has a stationary point when $\bar{k} = \dfrac{1-a}{a|b|}$. Regardless of the value of b, the second derivative given by (B10) is always negative. So, when $b > 0$, the positive first and negative second derivative implies that M is increasing and concave in $\bar{k}$, and as depicted in Figure B1, the steady state occurs at point A.

Figure B1: Depiction of the unique steady state when $b > 0$



When $b < 0$, the negative second derivative implies that the stationary point is a maximum. In this instance, if $M_{max} = \left[\frac{1-a}{a|b|}\right]^a a^{1-a} < \frac{\beta(1-a)}{1+\beta}$, there will be no non-trivial steady state, as depicted in Figure B2. Otherwise, there are two steady states, one to the left of the stationary point and one to the right. These steady state capital stocks, denoted by $\bar{k}_1^*$ and $\bar{k}_2^*$ are depicted graphically in Figure B3.

Figure B2: Depiction of a situation where there is no steady state when $b < 0$

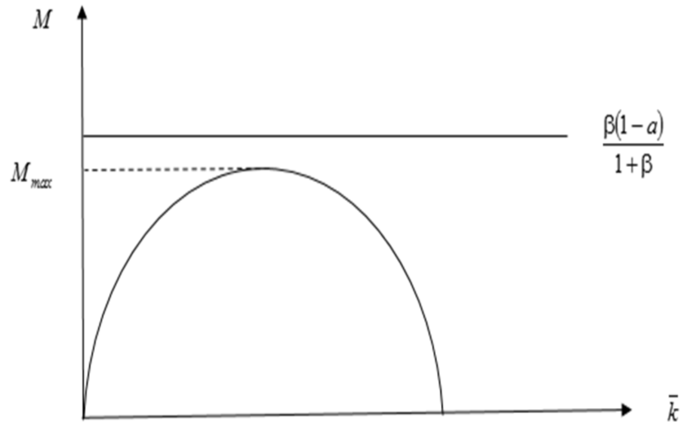
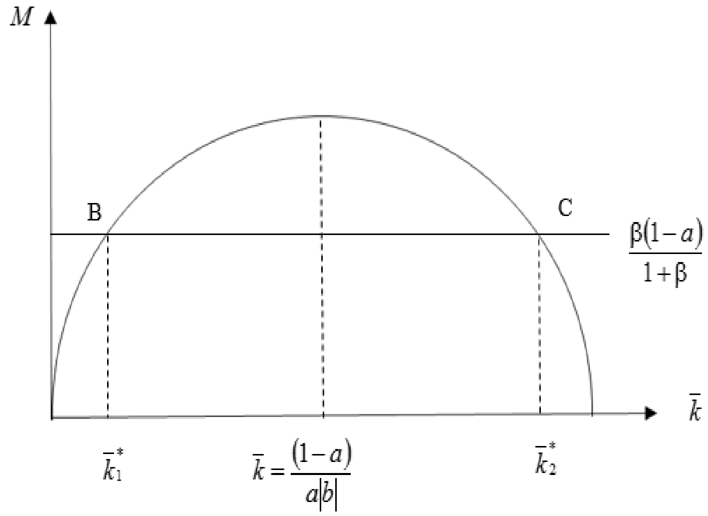


Figure B3: Depiction of the two possible steady states when $b < 0$



Recall that the condition necessary for stability of the steady state when $b < 0$ was given by $\bar{k} < \dfrac{(1-a)}{a|b|}$. So, evidently, $\bar{k}_1^*$ is a stable steady state while $\bar{k}_2^*$ is not.

Recall that the elasticity of substitution at any given steady state value must satisfy the condition $\bar{\sigma} = 1 + b\bar{k} > 0$. When $b > 0$, this condition is always satisfied. However, recall that when $b < 0$, the condition results in an upper bound for any steady state such that $\bar{k} < {}^{1}\!/\!_{|b|}$. This implies that the economy's capital stock can never rise above ${}^{1}\!/\!_{|b|}$. This also means that, although we are able to demonstrate that there may be two possible steady states, they can only exist if they are below this bound. Firstly, when $a > \frac{1}{2}$, $\frac{(1-a)}{a|b|} < \frac{1}{|b|}$. Since, $< \bar{k}_1^* < \frac{(1-a)}{a|b|} < \frac{1}{|b|}$, it clearly exists, and is a stable steady state. When $a > \frac{1}{2}$, $\bar{k}_2^*$ exists only if it falls in the range: $\frac{(1-a)}{a|b|} < \bar{k}_2^* < \frac{1}{|b|}$. If it falls within this range, if the economy starts with an initial capital stock such that $\bar{k}_2^* < k_0 < {}^{1}\!/\!_{|b|}$, capital stock will increase until it reaches ${}^{1}\!/\!_{|b|}$. On the other hand, if $\bar{k}_2^* > {}^{1}\!/\!_{|b|}$ it is non-existent as it violates the upper bound. Hence, we can conclude that at least one steady state exists when $a > \frac{1}{2}$.

Similarly, when $a < \frac{1}{2}$, $\frac{(1-a)}{a|b|} > \frac{1}{|b|}$. As $\bar{k}_2^* > \frac{(1-a)}{a|b|} > \frac{1}{|b|}$, it is clearly non-existent. When $a < \frac{1}{2}$, if $\bar{k}_1^* < {}^{1}\!/\!_{|b|}$, it is still the unique and stable steady state, but it is non-existent if it falls in the range $\frac{1}{|b|} < \bar{k}_1^* < \frac{(1-a)}{a|b|}$. Hence, when $a < \frac{1}{2}$, it may be possible that both steady states depicted in Figure B3 may not exist. Under such circumstances, the economy will regress towards the trivial steady state $\bar{k} = 0$. Hence, when $a < \frac{1}{2}$, the economy may fall into a poverty trap.

Appendix C

Proof of Proposition 2 in Chapter 3

Recall from equation (3.3.12) that the steady state is given by the following implicit function:

$$F(\bar{k}, b) = \bar{k}^{1-a}(1 + ab\bar{k})^a - \frac{\beta(1-a)}{(1+\beta)} = 0 \tag{C1}$$

From (C1) above we have

$$F_{\bar{k}} = (1-a)\bar{k}^{-a}(1 + ab\bar{k})^a + a^2 b\bar{k}^{1-a}(1 + ab\bar{k})^{a-1} = \left(\frac{1+ab\bar{k}}{\bar{k}}\right)^{a-1}\left(\frac{1-a+ab\bar{k}}{\bar{k}}\right) \tag{C2}$$

$$F_b = \left(\frac{1+ab\bar{k}}{\bar{k}}\right)^{a-1} a^2\bar{k} \tag{C3}$$

From the implicit differentiation theorem, we have:

$$\frac{\partial\bar{k}}{\partial b} = -\frac{F_b}{F_{\bar{k}}} = \frac{-a^2\bar{k}^2}{(1-a+ab\bar{k})} \tag{C4}$$

It is clear that, when $b > 0$, $\frac{\partial\bar{k}}{\partial b} < 0$. Recall that when $b > 0$, the unique steady state that emerges is always stable.

When $b < 0$, $\frac{\partial\bar{k}}{\partial b} < 0$ if $\bar{k} < \frac{(1-a)}{a|b|}$. From Proposition 1, this upper bound on $\bar{k}$ is also the condition necessary for a stable steady state. Thus, when $b < 0$, for any stable steady state, there is a negative relationship between $\bar{k}$ and b.

Thus, when $b<0,$ the conditions under which $\dfrac{\partial\bar{k}}{\partial b}<0$ are analogous to the conditions under which the steady state that emerges when $b<0$ is locally stable. Hence, when $b<0,$ any locally stable steady state capital stock is decreasing in the parameter b.

At steady state,

$$\bar{y}=\bar{k}^{a}\left(1+ab\bar{k}\right)^{1-a} \tag{C5}$$

$$\frac{d\bar{y}}{db}=a\left(\frac{1+ab\bar{k}}{\bar{k}}\right)^{1-a}\frac{d\bar{k}}{db}+(1-a)\left(\frac{1+ab\bar{k}}{\bar{k}}\right)^{-a}\left(a\bar{k}+ab\frac{d\bar{k}}{db}\right)=\left(\frac{1+ab\bar{k}}{\bar{k}}\right)^{-a}\left[a\frac{d\bar{k}}{db}\left(\frac{1+b\bar{k}}{\bar{k}}\right)+(1-a)a\bar{k}\right] \tag{C6}$$

The condition necessary for $d\bar{y}/db>0$ is therefore:

$$a\frac{d\bar{k}}{db}\left(\frac{1+b\bar{k}}{\bar{k}}\right)+(1-a)a\bar{k}>0 \tag{C7}$$

Substituting for $d\bar{k}/db$ and simplifying gives:

$$ab(1-2a)\bar{k}>2a-1 \tag{C8}$$

When $b>0,$ this yields:

$\bar{k}>\dfrac{-1}{ab}$, which is always satisfied because $\bar{k}>0$.

When $b < 0$, we get $\bar{k} < \frac{1}{a|b|}$, which is always satisfied for any feasible steady state. Hence, for any feasible steady state $\frac{d\bar{y}}{db} > 0$.

Appendix D

Optimal solutions to the agent's decision problem in Chapter 4

Under the assumption of log utility, (2.11) simplifies to:

$$c_{t+2} = \beta R_{t+2} c_{t+1} \tag{D1}$$

Using (2.1), (2.3) and (2.8), and assuming logarithmic preferences, we get:

$$\frac{\partial V(h_{t+2})}{\partial e^P_{t+1}} = \frac{1}{h_{t+2}} \times \frac{\gamma h_{t+2}}{e_{t+1}} \times \frac{(1-a)e_{t+1}}{e^P_{t+1} + abe^G_{t+1}} \tag{D2}$$

Simplifying (D2) and substituting into (2.12) yields:

$$\frac{\theta\gamma(1-a)}{e^P_{t+1} + ab\tau w_{t+1} h_{t+1}} = \frac{1}{c_{t+1}} \tag{D3}$$

(D3) can be rearranged to yield:

$$e^P_{t+1} = \theta\gamma(1-a)c_{t+1} - ab\tau w_{t+1} h_{t+1} \tag{D4}$$

Using equations (D1) and (D4), we can derive the optimal solutions to the agent's decision problem given by equations (4.3.13) to (4.3.16).

Appendix E

Proof of Proposition 1 in Chapter 4

From (4.2.13), we get:

$$U(c_{t+1}) = \ln\left[\frac{w_{t+1}h_{t+1}(1-\tau+ab\tau)}{1+\beta+\theta\gamma(1-a)}\right] \quad \text{(E1)}$$

Then,

$$\frac{\partial U(c_{t+1})}{\partial b} = \frac{a\tau}{1-\tau+ab\tau} \quad \text{(E2)}$$

Similarly, from (4.2.16) we get: $\dfrac{\partial U(c_{t+2})}{\partial b} = \dfrac{a\tau}{1-\tau+ab\tau}$.

In the case of human capital, substituting (4.2.7) and (4.2.14) into (4.2.1) and then substituting into (4.2.3) gives us:

$$U(h_{t+2}) = \ln\left\{h_{t+1}^{\delta}\left[w_{t+1}h_{t+1}\tau^{a}\left[\frac{\theta\gamma(1-a)(1-\tau+ab\tau)}{1+\beta+\theta\gamma(1-a)}\right]^{1-a}\right]^{\gamma}\right\}. \quad \text{(E3)}$$

Differentiating (E3) with respect to b gives us: $\dfrac{\partial U(h_{t+2})}{\partial b} = \dfrac{a\tau}{1-\tau+ab\tau}$.

As all three components of the utility function are increasing in b, the consumer's entire utility is increasing in b.

Appendix F

Derivation of the upper and lower bounds on *b*

First, note from (4.2.13), (4.2.15) and (4.2.16) that for c_{t+1}, s_{t+1} and c_{t+2} to be interior, the condition $1-\tau+ab\tau>0$ should hold. This simplifies to $b>m$, where $m=-\frac{(1-\tau)}{a\tau}$.

From (4.2.14), for $e_{t+1}^{P}>0$, we need $\theta\gamma(1-a)(1-\tau)-ab\tau(1+\beta)>0$. This condition simplifies to: $b<n$, where $n=\frac{\theta\gamma(1-a)(1-\tau)}{a\tau(1+\beta)}$

To get the modified upper bound for *b* that is needed to ensure that in any given period, the elasticity of substitution between public and private education expenditures is positive, we first substitute the expression for e_{t+1}^{G} from (4.2.7) and the optimal value for e_{t+1}^{P} given by (4.2.14) into (4.2.2) which yields the following inequality:

$$\eta=1+\frac{b\tau w_{t+1}h_{t+1}}{w_{t+1}h_{t+1}\left[\frac{\theta\gamma(1-a)(1-\tau)-ab\tau(1+\beta)}{1+\beta+\theta\gamma(1-a)}\right]}>0 \qquad \text{(F1)}$$

Noting that $\eta\geq 0$ and making *b* the subject of the resulting inequality yields the lower bound of $b>p$. However, note that in Section 2, the condition necessary for interior solutions for the optimal values of the decision variables was $b>\max\{-1,m\}$. Given $p=\frac{-\theta\gamma(1-\tau)}{\tau(1+\beta+\theta\gamma)}$ and $m=\frac{-(1-\tau)}{a\tau}$, p is binding if $p>m$. The required condition for $p>m$ is: $\theta\gamma a<1+\beta+\theta\gamma$, which is always satisfied since $0\leq a\leq 1$.

Therefore, the range of values for b yielding interior solutions *and* satisfying the condition $\eta > 0$ is $p < b < n$.

Appendix G

Derivation of the eigenvalues

The eigenvalues μ_1 and μ_2 are given by the following equations.

$$\mu_1 + \mu_2 = trA \tag{G1}$$

$$\mu_1\mu_2 = detA \tag{G2}$$

Upon substituting the expressions for the trace and the determinant of the Jacobian matrix A into (G1) and (G2), we get:

$$\mu_1 + \mu_2 = \beta j\lambda\left(\frac{\bar{k}}{\bar{h}}\right)^{\lambda-1} + l[\delta + (1-\lambda)\gamma]\bar{k}^{\lambda\gamma}\bar{h}^{-\delta+(1-\lambda)\gamma-1} \tag{G3}$$

$$\mu_1\mu_2 = \beta j\lambda l\delta\bar{k}^{\lambda(1+\gamma)-1}\bar{h}^{-\delta+\gamma-\lambda(1+\gamma)} \tag{G4}$$

By combining the expressions and setting $s_1 = \beta j\lambda\left(\frac{\bar{k}}{\bar{h}}\right)^{\lambda-1} + l[\delta + (1-\lambda)\gamma]\bar{k}^{\lambda\gamma}\bar{h}^{-\delta+(1-\lambda)\gamma-1}$ and $s_2 = \beta j\lambda l\delta\bar{k}^{\lambda(1+\gamma)-1}\bar{h}^{-\delta+\gamma-\lambda(1+\gamma)}$, we get the solutions for μ_1 and μ_2 given in (4.3.10) and (4.3.11).

Appendix H

Proof of proposition 3 in Chapter 4

To limit the possible characterisations of the steady state to cases (i) to (iii), we need to prove that $0 < s_1, s_2 < 1$.

Note that $j < 1$ if $b < \frac{1+\beta+\theta\gamma(1-a)+\tau-1}{a\tau}$. Given the upper bound on b, which was derived earlier to be $n = \frac{\theta\gamma(1-a)(1-\tau)}{a\tau(\beta+1)}$, this condition is always satisfied. Now, it is clear that all the coefficients of the expressions in (3.4) and (3.5) are below 1. Hence, $0 < \bar{k}, \bar{h} < 1$. As $0 < \bar{h} < 1$, it is clear from (3.6) that $0 < \frac{\bar{k}}{\bar{h}} < 1$. Hence, it follows that $0 < s_1, s_2 < 1$. This limits the number of possible characterisations of the steady state to cases (i) to (iii) outlined in Proposition 3.

Appendix I

Derivation of the sufficient condition for the emergence of case (i) of proposition 3 in Chapter 4

Expanding the term inside the square root in equation (4.3.10) or (4.3.11), we get:

$$\left\{\beta j\lambda\left(\frac{\bar{k}}{\bar{h}}\right)^{\lambda-1}+l[\delta+(1-\lambda)\gamma]\bar{k}^{\lambda\gamma}\,\bar{h}^{\delta+(1-\lambda)\gamma-1}\right\}^{2}-4\beta j\lambda l\delta\bar{k}^{\lambda(1+\gamma)-1}\bar{h}^{\delta+\gamma-\lambda(1+\gamma)} \quad \text{(I1)}$$

If we ignore the squared terms, the following is a sufficient condition to ensure that (I1) is positive:

$$2\beta j\lambda l[\delta+(1-\lambda)\gamma]\bar{k}^{\lambda(1+\gamma)-1}\bar{h}^{\delta+\gamma-\lambda(1+\gamma)}-4\beta j\lambda l\delta\bar{k}^{\lambda(1+\gamma)-1}\bar{h}^{\delta+\gamma-\lambda(1+\gamma)}>0 \quad \text{(I2)}$$

This simplifies to $\sigma<(1-\lambda)\gamma$.

Appendix J

Derivation of the optimal value of τ in Chapter 4

Let I be the indirect utility function.

$$I = U\left(c_{t+1}^{*}\right) + \beta U\left(c_{t+2}^{*}\right) + \theta V\left(h_{t+2}^{*}\right) \tag{J1}$$

From (4.3.13), (4.3.14) and (4.3.16) we have:

$$\frac{\partial U\left(c_{t+1}^{*}\right)}{\partial \tau} = \frac{\partial U\left(c_{t+2}^{*}\right)}{\partial \tau} = \frac{ab-1}{1-\tau+ab\tau} \tag{J2}$$

$$\frac{V\left(h_{t+2}^{*}\right)}{\partial \tau} = \theta\gamma\left[\frac{a}{\tau} + \frac{(1-a)(ab-1)}{1-\tau+ab\tau}\right] \tag{J3}$$

Substituting these expressions into the derivative of (J1), we can obtain the optimal value τ^{*} given by equation (4.4.1).

The second derivatives associated with (J2) and (J3) are negative, which confirms that at τ^{*} indirect utility is maximised.

Appendix K

Derivation of the critical wealth level $\hat{W}_t$ in Chapter 5

If $W_t \le \hat{W}_t$, $h_t^P \le \hat{h}_t^P$.

Substituting for h_t^P and $\hat{h}_t^P$ in the second inequality above yields:

$$\frac{Y_t - \gamma ab\psi\tau\overline{W_t}}{(1+\gamma)} \le \left[\frac{\left(\frac{\overline{\phi}}{x}\right)^{\frac{1}{\varepsilon}}}{A\left(\psi\tau\overline{W_t}\right)^{a}}\right]^{\frac{1}{1-a}} - ab\psi\tau\overline{W_t} \tag{K1}$$

Substituting for Y_t and rearranging yields:

$$\hat{W}_t = \frac{\left[\frac{\left(\frac{\overline{\phi}}{x}\right)^{\frac{1}{\varepsilon}}}{A\left(\psi\tau\overline{W_t}\right)^{a}}\right]^{\frac{1}{1-a}}(1+\gamma) - \tau\overline{W_t}\left[(ab-1)\psi+1\right] - w_t - \tau(1-\psi)\overline{W_t} - \frac{w_2}{R}}{(1-\tau)} \tag{K2}$$

Thus, any agent with wealth (received as intended or unintended bequests) greater than $\hat{W}_t$ can afford the critical private health expenditure $\hat{h}_t^P$.

Appendix L

Proof of Bimodality in Chapter 5

In any period t, the offspring's wealth is equal to the intended bequest such that $W_t = m_t$ if the parent survives through old age, and the offspring's wealth is equal to the parent's savings such that $W_t = s_{t-1}$ if the parent dies prematurely. At steady state, $W_{t+1} = W_t = W^*$.

First, using the characterisation of intended bequests given by equation (5.3.14), for agents with wealth levels $W_t \leq \hat{W}_t$, steady state wealth is:

$$W^* = \frac{\gamma \overline{R}\left[\overline{w}_1 + (1-\psi+ab\psi)\tau\overline{W} + \frac{\overline{w}_2}{\overline{R}}\right]}{(\beta+1)(\gamma+1) - \gamma\overline{R}(1-\tau)} \tag{L1}$$

For agents with wealth levels $W_t > \hat{W}_t$, steady state wealth is given by:

$$W^* = \frac{\overline{w}_1 + (1-\psi)\tau\overline{W} + \frac{\overline{w}_2}{\overline{R}} - \hat{h}_t^p}{(\beta+1)\overline{R} - (1-\tau)} \tag{L2}$$

In the case of unintended bequests, using equation (3.15), for agents with wealth levels $W_t \leq \hat{W}_t$, steady state wealth is:

$$W^* = \frac{\gamma\left[\overline{w}_1 + (1-\psi+ab\psi)\tau\overline{W} + \frac{\overline{w}_2}{\overline{R}}\right] - \frac{\overline{w}_2}{\overline{R}}}{1+\gamma\tau} \tag{L3}$$

For agents with wealth levels $W_t > \hat{W}_t$, steady state wealth is given by:

$$W^* = \frac{\overline{w}_1 + (1-\psi)\tau\overline{W} - \hat{h}_t^p}{(\beta+1)\overline{R} - (1-\tau)} \qquad \text{(L4)}$$

For these steady states to be stable, it is necessary that the intended and unintended bequest functions for both types of agents in the economy should have slopes less than 1. To demonstrate that this is indeed the case, first we consider the slope of the intended bequest line for agents with wealth levels $W_t \le \hat{W}_t$, which, from (5.3.10) is given by:

$$\frac{(1-\tau)\gamma\overline{R}}{(\beta+1)(1+\gamma)} \qquad \text{(L5)}$$

Recall that the condition $\overline{R} > (\beta+1)$ was imposed to ensure that intended bequests are higher than unintended bequests for all agents. Thus, for $0 < \frac{(1-\tau)\gamma\overline{R}}{(\beta+1)(1+\gamma)} < 1$, we need $0 < \frac{\overline{R}}{(\beta+1)} < \frac{(1+\gamma)}{(1-\tau)\gamma}$. As we know that $\frac{\overline{R}}{(\beta+1)} > 1$, it follows that the condition necessary for the slope to be below 1 is $\frac{(1+\gamma)}{(1-\tau)\gamma} > 1$, which is always satisfied.

For agents with wealth levels $W_t > \hat{W}_t$, from (5.3.10), the slope of the intended bequests line is:

$$\frac{\overline{R}(1-\tau)}{(\beta+1)} \qquad \text{(L6)}$$

Given that $\frac{\overline{R}}{(\beta+1)} > 1$, the condition $0 < \frac{\overline{R}(1-\tau)}{(\beta+1)} < 1$ is satisfied as long as $\frac{\overline{R}}{(\beta+1)} < \frac{1}{1-\tau}$, which is indeed the case, since the ratio $\frac{\overline{R}}{(\beta+1)}$ is only marginally higher than 1.

In the case of unintended bequests, it can be seen from (5.3.11) that for agents with wealth levels $W_t \leq \hat{W}_t$, the slope is given by $\frac{\gamma(1-\tau)}{(1+\gamma)}$, which is clearly always below 1. For agents with wealth levels $W_t > \hat{W}_t$, from (5.3.11), the slope of the unintended bequests line is simply equal to $(1-\tau)$, which is clearly less than 1.

Appendix M

Proof of Propositions 3 and 4 in Chapter 5

Proposition 3: For agents with $W_t \leq \hat{W}_t$, we derive ψ^* by obtaining the indirect utility function which we get by substituting optimal values of ϕ, c_{t+1} and m^*_{t+1} into (3.1), which yields:

$$\phi^*(\psi)\left\{U\left[c^*_{t+1}(\psi)\right]+\theta V\left[m^*_{t+1}(\psi)\right]\right\} \tag{M1}$$

Differentiating (M1) yields the FOC:

$$\frac{d\phi^*}{dh^*_t}\times\frac{dh^*_t}{d\psi}\left\{U\left[c^*_{t+1}(\psi)\right]+V\left[m^*_{t+1}(\psi)\right]\right\}+\phi^*(\psi)\left\{\frac{dU\left(c^*_{t+1}\right)}{dc^*_{t+1}}\times\frac{dc^*_{t+1}}{d\psi}+\frac{dV\left(m^*_{t+1}\right)}{dm^*_{t+1}}\times\frac{dm^*_{t+1}}{d\psi}\right\}=0 \tag{M2}$$

Noting that:

$$h^*_t=\left(\psi\tau\overline{W}\right)^a\left(h^{P^*}_t+ab\psi\tau\overline{W}\right)^{1-a} \tag{M3}$$

$$\frac{\partial h^*_t}{\partial\psi}=h^*_t\left[\frac{a}{\psi}+\frac{\tau\overline{W}(1-a)(ab-1)}{\left(Y_t+ab\psi\tau\overline{W}\right)}\right] \tag{M4}$$

We get:

$$\frac{\dfrac{d\phi}{dh_t}\times\dfrac{dh_t}{d\psi}}{\phi^*(\psi)}=\varepsilon\left[\frac{a}{\psi}+\frac{\tau\overline{W}(1-a)(ab-1)}{\left(Y_t+ab\psi\tau\overline{W}\right)}\right] \tag{M5}$$

Further,

$$U\left[c^*_{t+1}(\psi)\right]+\theta V\left[m^*_{t+1}(\psi)\right]=m^{*\,1-\sigma}_{t+1}\left(\frac{\beta^{1-\sigma}+\theta}{1-\sigma}\right) \tag{M6}$$

and,

$$\frac{dU\left(c_{t+1}^*\right)}{dc_{t+1}^*}\times\frac{dc_{t+1}^*}{d\psi}+\frac{dV\left(m_{t+1}^*\right)}{dm_{t+1}^*}\times\frac{dm_{t+1}^*}{d\psi}=\frac{{m_{t+1}^*}^{-\sigma}\gamma\overline{R}\tau\overline{W}(ab-1)\left(\beta^{1-\sigma}+\theta\right)}{(1+\beta)(1+\gamma)} \quad \text{(M7)}$$

Substituting these expressions into (M2) and substituting the functional form for b_{t+1}^* yields:

$$\varepsilon\left[\frac{a}{\psi}+\frac{\tau\overline{W}(1-a)(ab-1)}{\left(Y_t+ab\psi\tau\overline{W_t}\right)}\right]\times\left(Y_t+ab\psi\tau\overline{W_t}\right)=-\tau\overline{W_t}(ab-1)(1-\sigma) \qquad \text{(M8)}$$

Substituting for Y_t and rearranging the above expression yields the optimal solution:

$$\psi^*=\frac{\varepsilon a\left[\overline{w}_1+(1-\tau)W_t+\tau\overline{W_t}+\dfrac{\overline{w}_2}{\overline{R}}\right]}{\tau\overline{W_t}(1-ab)(1+\varepsilon-\sigma)} \qquad \text{(M9)}$$

The second derivative is:

$$\frac{-\varepsilon a\left[w_t+(1-\tau)W_t+\tau\overline{W_t}+\dfrac{\overline{w}_2}{\overline{R}}\right]}{\psi^2} \qquad \text{(M10)}$$

This expression is always negative, which confirms that at ψ^* the indirect utility function reaches its maximum value.

Given $0<\sigma<1$ and $\varepsilon<1$, $1+\varepsilon-\sigma>0$. So, for ψ^* to be positive, it is necessary that: $1-ab>0$ which yields $b<\frac{1}{a}$. However, for an interior solution (i.e. $0<\psi^*<1$), the required condition is:

$$\varepsilon a\left[\overline{w}_1+(1-\tau)W_t+\tau\overline{W_t}+\frac{\overline{w}_2}{\overline{R}}\right]<-\tau\overline{W}(ab-1)(1+\varepsilon-\sigma),$$

which yields the following upper bound for *b*:

$$b < \frac{1}{a} - \frac{\varepsilon\left[\overline{w}_1 + (1-\tau)W_t + \tau\overline{W_t} + \frac{\overline{w}_2}{\overline{R}}\right]}{\tau\overline{W_t}(1+\varepsilon-\sigma)}.$$

For agents with $W_t > \hat{W}_t$, the first order condition is:

$$\overline{\phi}\left[\frac{dU(c^*_{t+1})}{dc^*_{t+1}} \times \frac{dc^*_{t+1}}{d\psi} + \frac{\theta dV(m^*_{t+1})}{dm^*_{t+1}} \times \frac{dm^*_{t+1}}{d\psi}\right] = 0 \qquad \text{(M11)}$$

Substituting for the expressions in (M7) and noting that:

$$\frac{\partial \hat{h}^p_t}{\partial \psi} = -\left(\frac{\overline{\phi}}{x}\right)^{1/\varepsilon(1-a)} \frac{a}{(1-a)\left(\tau\overline{W_t}\right)^a \psi^{1/1-a}} - ab\tau\overline{W_t} \qquad \text{(M12)}$$

We obtain the derivative of the indirect utility function with respect to ψ:

$$\frac{\partial U^*}{\partial \psi} = \left(\beta^{1-\sigma} + \theta\right)\left\{\frac{w_{t+1} + \overline{R}\left[w_t + (1-\tau)W_t + \tau(1-\psi)\overline{W_t} - \hat{h}^p_t\right]}{1+\beta}\right\}^{-\sigma} \times \overline{R}\left[-\tau\overline{W_t} + \left(\frac{\overline{\phi}}{x}\right)^{1/\varepsilon(1-a)} \frac{a}{(1-a)\left(\tau\overline{W_t}\right)^a \psi^{1/(1-a)}} + ab\tau\overline{W_t}\right] = 0$$

(M13)

From (M13), we get:

$$\psi^* = \left(\frac{\overline{\phi}}{x}\right)^{\frac{1}{\varepsilon}}\left[\frac{a}{(1-a)(1-ab)\left(\tau\overline{W_t}\right)^{1+a}}\right]^{1-a} \qquad \text{(M14)}$$

If we assume $q = \frac{1}{a} - \frac{\varepsilon\left[\overline{w}_1 + (1-\tau)W_t + \tau\overline{W_t} + \frac{\overline{w}_2}{\overline{R}}\right]}{\tau\overline{W_t}(1+\varepsilon-\sigma)}$, the optimal solutions can be characterised as in equation (5.3.14).

Proposition 4: From equations (M9) and (M14) it is clear that ψ^* is rising in b for all agents in the economy